Evidence-Based Incentive Systems with an Application in Health Care Delivery

Donald K.K. Lee
Department of Management Science & Engineering, Stanford University, Stanford CA 94305
Stefanos A. Zenios
Graduate School of Business, Stanford University, Stanford CA 94305

This paper develops an empirical method to estimate the parameters of a multi-task principal-agent model. The principal has a stake in the performance of a system, but delegates its control to an agent. The agent chooses which tasks to perform and the effort to put in each task. The principal observes the system’s aggregate performance (downstream outcome) and several other performance measures (typically process-compliance measures to be referred to as intermediate outcomes). All observed measures are noisy signals of the agent’s effort in each task. The principal rewards the agent based on a weighted combination of the observed performance measures. The question is to determine the optimal mix of performance measures that would maximize the principal’s expected payoff. Using the Empirical Likelihood method, we show how the principal can use data from multiple agents to answer the following questions: a) How can intermediate process-compliance measures be integrated into a single intermediate performance score that can be used in an optimal payment system? b) What is the agent’s cost of effort and reservation utility? c) What is the optimal payment system? The method was applied to data from patients with kidney failure who needed dialysis (Medicare, the payer, was the principal and the dialysis providers were the agents). An optimal payment system was designed. The system was shown to have the potential to increase the number of hospital-free days per patient year-at-risk by 7.4% without increasing total medical expenses.

Key words: Principal-Agent Models; Empirical Likelihood Method; Moment Conditions.

History: First Version: September 2007

1. Introduction

The following question is common in the design of incentive systems for individuals or organizations: What is the appropriate mix of performance measures that will induce optimal behavior? The multi-task principal-agent paradigm has proven useful in addressing this question and providing stylized insights. In this paradigm, the principal has a stake in the performance of a system but
delegates its control to an agent. The agent chooses which tasks to perform and the effort to put in each task. The principal observes the system’s aggregate performance (to be referred to as the downstream outcome) and several other performance measures (typically process-compliance measures to be referred to as intermediate outcomes) which are noisy indicators of the agent’s effort in each task. The principal rewards the agent based on a weighted combination of the observed performance measures. The question is to determine the optimal mix of performance measures that would maximize the principal’s expected payoff.

Gibbons (4) provides a review of the literature on principal-agent models and places it in the context of problems of interest to Operations Management scholars. The following important insight is highlighted in that review: An intermediate performance measure is valuable, and should be used in a performance payment system, if it is aligned with the principal’s interest. Its inclusion in the payment system would induce useful actions that would maximize the principal’s payoff. But then this motivates the following question: How can one empirically identify intermediate performance measures that are aligned with the principal’s interests? The main contribution of this paper is to propose an empirical methodology to identify appropriate intermediate performance measures that would induce valuable actions, estimate the cost of these actions, and embed these estimates in a multi-task principal-agent model in order to design optimal payment systems. The methodology is developed in the context of the payment systems for the treatment of patients with End-Stage Renal Disease (ESRD, also known as terminal kidney failure).

To motivate the methodology, we start with a stylized multi-task principal-agent model which we use to address the following question: When should the principal use intermediate outcomes in combination with the downstream outcome? The answer is intuitive: if the effort that affects intermediate outcomes also affects downstream outcomes, then incorporating the intermediate outcome is advantageous if its inclusion reduces the variability in the agent’s payment, and thus reduces the risk premium the principal will need to pay the agent to persuade him to provide the desired services. Section 2 presents a model that articulates this intuition.
We then proceed to develop an empirical method to identify intermediate performance measures that have this property, and then estimate the parameters of the principal-agent model in order to design the optimal payment system. This is done using data from ESRD. Section 3 modifies the conceptual principal-agent model to fit the characteristics of payer (principal) and providers (agents) of ESRD services. The estimation method is described in section 4. The method consists of two steps: First, it estimates the functional form of the intermediate and downstream outcomes to include in the payment system and second, the agent’s cost of effort. Armed with these parameters, it is then possible to obtain the optimal payment system. The ESRD data are described in section 5 and the results from the estimation method presented in section 6. Concluding remarks appear in Section 7. The online companion provides background information on the empirical methods, summary statistics for the data used in the paper, and examples of other applications of the empirical methodology.

Motivating Example: Medicare’s End-Stage Renal Disease (ESRD) Program. We will now provide background information on ESRD in order to further motivate the contributions of this manuscript.

The ESRD program was established in 1973 by US Congress to cover the medical costs of individuals afflicted with kidney failure (ESRD). The program currently pays for the healthcare costs of about 85% of ESRD patients in the US ($16 billion in 2003 and growing at a 10% annual rate). The majority of these patients require dialysis therapy three times a week at a dialysis center, and the dialysis provider (agent) is reimbursed by Medicare (principal).

The Dialysis Process. During the process of dialysis, the patient’s blood flow is connected to the dialysis machine through an access point on his or her hand (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. Adequate dialysis requires sufficiently long treatment in order to effectively remove the majority of the toxins. The amount of toxins removed during each treatment is measured by the so-called Urea Reduction Ratio (URR) and existing medical guidelines suggest a URR of at least 65%. A complicated formula is used by the dialysis provider to determine the appropriate
duration of the treatment that would achieve the desired URR. The formula takes into account the patient’s body size and weight, the blood flow rate though the access point, and the efficiency of the dialysis machine. For the typical patient, a treatment between three and four hours achieves a URR of 65%. During each dialysis treatment, the provider also administers several injectable drugs to rectify biological imbalances caused by kidney failure. The most important drug is Erythropoietin Alpha (EPO) to treat the anemia caused by the failing kidneys interference with the normal regulation of red blood cells. Clinical guidelines require that a patient’s hematocrit level (a measure of red blood count) be within the optimal range of 33% to 36%. By dedicating adequate resources (intermediate effort) to the management of URR and anemia (intermediate outcomes), each provider can achieve outcomes consistent with the clinical guidelines. Specific activities (intermediate tasks) that are known to enhance these outcomes include regular monitoring of the patient’s blood flow in the access point, frequent measurement of patient’s weight, body size and hematocrit levels, quality control systems for dialysis filters, etc. While these intermediate activities may be costly for the provider, they enhance patient’s URR and hematocrit level.

The complications that arise from the dialysis treatment itself may cause patients to be hospitalized (downstream outcome). Providers influence hospital admissions rate by ensuring that the URR and hematocrit are in the ranges recommended by clinical guidelines, but also by engaging in specific hospital-prevention strategies. 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 4 lbs in fluids between treatments have higher hospital admission rates. The dialysis clinics, however, can reduce these admissions rate by implementing various measures to manage excessive weight gain between treatments: dietary counseling, performing an additional dialysis treatment for patients with a history of excessive weight gain, equipping each heart failure patient with a scale connected to the internet that the patient uses to monitor his or her weight gain and which can notify his or her physician when excessive weight gain indicates
elevated risk for heart failure. Such downstream efforts, while costly to the provider, can have a significant impact in patient hospital admission rates.

Costs and Incentives. The costs for hospital admissions account for a third of the Medicare program’s expenditures. Therefore, it is in Medicare’s interest to maximize the amount of time its patients are hospital-free. Providers are currently paid on a fee-for-service basis, i.e. a fixed payment for each dialysis treatment performed plus an additional payment for any drugs administered in each treatment. This payment system creates partial incentives for the provider to reduce hospitalizations because of the opportunity cost of missed treatment when patients are hospitalized. But it is not clear if this is sufficient to fully align a profit-maximizing provider’s goals with Medicare’s objective of maximizing hospital-free days (downstream outcome).

In 2003, the US congress passed legislation which mandated the development of a pay-for-performance reimbursement system for Medicare’s ESRD programs (Medicare Modernization Act (12)). In this system, providers whose patients achieve superior downstream outcomes will be compensated more. In addition Medicare was asked to commission pilot studies to examine the feasibility of also rewarding dialysis providers for achieving a high fraction of the patients with adequate URR and optimal hematocrit levels.

Multi-task Principal-Agent Framing. In summary, Medicare (the principal) wishes to design a performance incentive system to motivate the dialysis providers (its agents) to minimize total hospital expenditures (the downstream outcomes). The payment system will combine intermediate outcomes based on dialysis dosage and anemia management and the downstream outcome of hospital admission rates. Processes and activities that agents take to control URR and Anemia represent the intermediate effort and they influence hospital admission rates. Any additional tasks that the providers perform in order to avert hospital admissions (e.g. weight gain management programs for patients with heart failure) reflect downstream effort and affect downstream outcomes.

In this paper, we will use the principal-agent paradigm to address three questions regarding the design of an optimal pay-for-performance system for ESRD. First, the observed intermediate and downstream outcomes cannot be used as performance metrics for rewarding providers without
certain adjustments. A patient with pre-existing health conditions (e.g., diabetes) is more likely to be hospitalized regardless of the provider’s effort. Therefore, outcomes will need to be controlled for each patient’s known health conditions (a procedure known as risk-adjustment in the medical community). Using a comprehensive list of patient attributes we show in section 4 how to obtain risk-adjusted intermediate and downstream outcomes to use as a basis for reimbursing providers.

The second question is how much should Medicare pay a provider that gave 80% of its patients adequate dosage and 60% of them sufficient anemia management, versus another provider that supplied 60% of its patients with adequate dosage and 80% sufficient anemia treatment? We use in section 4 a tensor product of two piecewise linear functions to construct a scalar function of the two variables that represents an intermediate outcome score for every combination of results. The intermediate score is designed to maximize its alignment with the portion of the provider’s intermediate effort that affect its downstream outcomes. This intermediate score then becomes the single intermediate outcome used in the performance payment system.

The third question is concerned with the design of the optimal payment system. To do this using the principal-agent framework, one needs to know: a) the provider’s cost of effort; b) the provider’s degree of risk aversion, and c) the covariance structure between the intermediate and downstream outcomes. We identify in section 4 statistical conditions to estimate a) and c) using patient observations from multiple dialysis providers. However, because the principal-agent framework considered here is not rich enough, it does not provide a statistical condition for estimating risk aversion. Nevertheless, the results in section 6 show that the optimal payment rates are insensitive to changes in the provider’s risk aversion and they can increase each patient’s expected hospital-free days by almost one month per year-at-risk without increasing Medicare’s budget or reducing provider profits.

Contributions to Literature. In the economics literature many reimbursement systems have been proposed for the principal-agent paradigm, but none of these can be readily implemented in practice. The limited empirical work that exists on the principal-agent models is usually concerned with examining whether the principal-agent theory explained the observed behavior in historical
data (Garen (3)). To the best of our knowledge, our work is the first attempt at integrating data-based parameter estimation with operationalization of a reimbursement contract.

Research pertaining to principal-agent models comes in two flavors. The first deals with hidden information, where the agent’s cost function is unknown to the principal and the principal's goal is to induce the agent to reveal its production costs. The agent is assumed to be either risk-neutral or have a pre-specified concave utility function. See for example the standard model studied in (Shleifer (18)), who showed in theory that it is optimal to reimburse an agent using the average of costs achieved by other agents. However from a practical viewpoint it is appealing to be able to directly estimate this cost function from empirical data. Working on the assumption that providers are profit-maximizing under the current reimbursement system, we show how to construct a local quadratic approximation to the agent’s cost function.

The second class of models analyze situations where the agent’s actions are unobservable by the principal (hidden action). The most relevant work to our study in the management science literature is (Fuloria & Zenios (2)), which analyzed a pay-for-performance scheme for dialysis before the enactment of Congress’ Medicare directive in 2003. The authors proposed a system under which Medicare payments to a dialysis provider is tied to patient hospitalizations and deaths. The model employed was highly stylized, and was an application of a dynamic principal-agent model with multiple time epochs. At the beginning of each epoch, the provider was paid a fixed fee for each patient it would treat during that epoch. At the end of the epoch, patient hospitalizations and/or deaths were observed and Medicare imposed a retrospective penalty on each adverse outcome, payable by the provider. The payment rates and penalties were then set to maximize Medicare’s expected discounted utility. Simulation results suggested that pay-for-performance schemes could in theory bring significant improvements to patient outcomes compared to the prevailing practice (at least in theory). However because the model invoked numerous simplifying assumptions, the authors advised against implementing their system. Also, that work came before Medicare’s recent initiatives to incorporate intermediate measures into its reimbursement algorithm, and hence it
did not consider these measures in the model. As a remedy we construct an evidence-based reimbursement system that makes efficient use of information from both intermediate and downstream outcomes. Because our system is data-driven and thus makes fewer parametric assumptions than (Fuloria & Zenios (2)), this study brings us one step closer to an actionable pay-for-performance reimbursement system for dialysis providers.

2. Conceptual Principal-Agent Model

We present a stylized principal-agent model with the goal of identifying conditions under which the inclusion of intermediate outcomes would be beneficial to the principal. The model is a special case of the multi-task model of (Holmstrom & Milgrom (6)) as reviewed in Gibbons (4). The standard assumptions that enable tractability are made: exponential utility for the agent, linear payment, and quadratic cost of effort.

The agent chooses a pair of actions (effort levels) \( e = (e_{int}, e_{ds}) \) that affect output: \( e_{int} \) is the component of the agent’s effort that influences both the intermediate outcome, \( INT_{out} \), and the downstream outcome \( DS_{out} \), while \( e_{ds} \) represents additional effort that affect \( DS_{out} \) alone. The agent’s production function is noisy: the agent’s efforts only partially determine the realization of \( INT_{out} \) and \( DS_{out} \) and the following linear relationship is assumed between efforts and outcomes,

\[
\begin{align*}
INT_{out} &= e_{int} + \varepsilon_{int} \\
DS_{out} &= e_{int} + e_{ds} + \varepsilon_{ds},
\end{align*}
\]

where \( \varepsilon = (\varepsilon_{int}, \varepsilon_{ds}) \) represent noise that affect outputs but is beyond the agent’s control. The noise terms have zero mean and covariance:

\[
\Sigma \equiv \text{Var}(\varepsilon) = \begin{pmatrix}
\sigma_{int}^2 & \sigma_{int,ds} \\
\sigma_{int,ds} & \sigma_{ds}^2
\end{pmatrix}
\]

The agent’s cost of effort is \( g(e_{int}, e_{ds}) = \frac{1}{2}e^T Q e \), his utility is \( U(x) = -e^{-rx} \) and the noise in (1) normal. The principal receives a reward \( \nu \) for each unit of downstream outcome, but not for the intermediate outcomes. She then compensates the agent according to a linear contract

\[
s(INT_{out}, DS_{out}) = \pi_0 + \pi_{int} INT_{out} + \pi_{ds} DS_{out}
\]
The principal’s problem is to choose the parameters of the payment system \( \pi = (\pi_0, \pi_{\text{int}}, \pi_{\text{ds}}) \) and the so-called vector of induced effort \( e = (e_{\text{int}}, e_{\text{ds}}) \) to maximize her expected reward:

\[
\nu E[DSout] - \pi_0 - \pi_{\text{int}} E[INTout] - \pi_{\text{ds}} E[DSout]
\]

subject to the following two constraints: a) The incentive compatibility constraint that the agent’s certainty equivalent is larger under the induced effort level \( e \) than any other effort level \( e' = (e'_{\text{int}}, e'_{\text{ds}}) \),

\[
\pi_0 + (\pi_{\text{int}} + \pi_{\text{ds}}) e_{\text{int}} + \pi_{\text{ds}} e_{\text{ds}} \geq g(e_{\text{int}}, e_{\text{ds}}) - \frac{r}{2} (\pi_{\text{int}}^2 \sigma_{\text{int}}^2 + 2\pi_{\text{int}} \pi_{\text{ds}} \sigma_{\text{int,ds}} + \pi_{\text{ds}}^2 \sigma_{\text{ds}}^2)
\]

b) the participation (or individual rationality) constraint that the agent’s optimal certainty equivalent exceeds a reservation utility \( U_0 \),

\[
\pi_0 + (\pi_{\text{int}} + \pi_{\text{ds}}) e_{\text{int}} + \pi_{\text{ds}} e_{\text{ds}} \geq U_0.
\]

The analysis of this model follows the method described in Gibbons (4) and leads into closed form expressions for the optimal payment system (see online companion, Table EC.1). Here, we will focus on: a) Comparing the optimal system to a system that utilizes either only the intermediate or only the downstream outcomes; b) Developing a method to calibrate and operationalize this model.

For the former, we consider a special case in which the matrices \( \Sigma \) and \( Q \) are diagonal, and \( Q_{11} = Q_{22} \). The following result quantifies the efficiency gain from incorporating the intermediate outcome when the noise in that outcome is small compared to that of the downstream outcome, i.e. \( \delta_\sigma \equiv \sigma_{\text{int}}^2 / \sigma_{\text{ds}}^2 \approx 0 \). (The proof uses a simple Taylor expansion and is omitted).

**Proposition 1.** Let \( \delta_\sigma \equiv \sigma_{\text{int}}^2 / \sigma_{\text{ds}}^2 \). 1) In the system that rewards both intermediate and downstream outcomes, the ratio of the optimal payments for downstream to intermediate outcomes is

\[
\frac{\pi_{\text{ds}}^*}{\pi_{\text{int}}^*} = \frac{1}{rQ_{11} \sigma_{\text{ds}}^2} + 2\delta_\sigma
\]
2) The principal’s expected additional gain from using a reimbursement system that rewards both outcomes compared to a system that only rewards downstream outcome is

\[
\frac{r \nu^2 \sigma_d^2}{2(2 + 3Q_{11} r \sigma_d^2 + Q_{11} r^2 \sigma_d^4)} \left( 16 + 18Q_{11} r \sigma_d^2 + 5Q_{11} r^2 \sigma_d^4 + Q_{11} r^3 \sigma_d^6 \right) - \frac{Q_{11} r^2 \nu^2 \sigma_d^4}{2(1 + Q_{11} r \sigma_d^2)^3} \left( 8 + 5Q_{11} r \sigma_d^2 + Q_{11} r^2 \sigma_d^4 \right) \delta_\sigma + O(\delta_\sigma^2)
\]

(8)

3) The additional gain over a system that only rewards intermediate outcome is

\[
\frac{2 \nu^2}{Q_{11}} \left( 2 + 5Q_{11} r \sigma_d^2 + Q_{11} r^2 \sigma_d^4 \right) - \frac{Q_{11} r^2 \nu^2 \sigma_d^4}{(1 + Q_{11} r \sigma_d^2)^2} \left( 6 + Q_{11} r \sigma_d^2 \right) \delta_\sigma + O(\delta_\sigma^2)
\]

(9)

The intuition is as follows: The value of incorporating the intermediate outcome increases as the ratio of the noise in the downstream outcome over the noise in the intermediate outcome increases. In addition, the coefficients in the optimal payment system are such that the weight assigned to the intermediate outcome increases as the marginal cost of effort, coefficient or risk aversion, or variance in the downstream outcome increases. But this raises the following practical consideration: How do we measure the noise ratio or the marginal cost of effort? Without knowledge of these quantities, the insights from the principal-agent model cannot be translated into precise payment systems. Our next step is to devise a systematic method for extracting these key parameters from data. Using data from ESRD patients, we devote subsequent sections to the empirical design of a pay-for-performance reimbursement system for Medicare’s ESRD program.

3. Adaptation to Dialysis Delivery

We now make two modifications to the conceptual model from section 2 to fit the Medicare dialysis program: a) The agent treats multiple patients with one intermediate and one downstream outcome for each patient; b) The outcomes are affected not only by the agent’s efforts but also by the patient’s characteristics so they will need to be adjusted to reflect these characteristics. The principal will design a linear payment system in which the provider’s compensation will be tied to the intermediate and downstream outcomes of its patients.

Consider a dialysis provider that treats \(N\) patients, and let \(INT_{out_i}\) and \(DS_{out_i}\) denote patient \(i\)’s observed intermediate and downstream outcomes respectively. \(DS_{out_i} \in [0, 1]\) is the proportion
of time patient $i$ is alive and not hospitalized. $\text{INT}_i$ can be a function of several intermediate quality measures. In practice, there are several intermediate quality measures that can be used to construct $\text{INT}_i$ but the following two described in the introduction are the most important: $\text{DOSAGE}_i$ is the proportion of patient $i$’s treatments with the blood urea reduction ratio exceeding a minimum standard of 65% (Urea Reduction Ratio (URR) ≥65%), and $\text{ANEMIA}_i$ is the proportion of patient $i$’s treatments with hematocrit level between the optimal range 33% to 36%. The intermediate outcome $\text{INT}_i$ can be viewed as a sufficient statistic that captures all the ‘relevant’ information contained in $(\text{DOSAGE}_i, \text{ANEMIA}_i)$:

$$\text{INT}_i = f(\text{DOSAGE}_i, \text{ANEMIA}_i)$$ (10)

where $f(\cdot, \cdot) : \mathbb{R}^2_+ \rightarrow \mathbb{R}$ remains to be determined. If Medicare is to pay providers for each unit of $\text{INT}$ ‘produced’, then $\text{INT}$ should be as closely aligned to the downstream outcome as possible. In other words $f(\cdot, \cdot)$ should be such that $\text{INT}$ equals the conditional expected downstream outcome $\mathbb{E}[\text{DSout}|\text{DOSAGE}, \text{ANEMIA}]$, a scalar quantity. As discussed in (Gibbons (4)), strong alignment between the principal’s objectives and the agent’s incentives allows for more effective contracts.

The outcomes $\text{INT}_i$ and $\text{DSout}_i$ will need to be risk-adjusted for patient $i$’s characteristics. Let $\text{PAT}_i$ denote a set of patient characteristics that affect observed patient outcomes (we will assume $\text{PAT}_i \in \mathbb{R}^K$). Then, the patient-level version of (1) is

$$\begin{align*}
\text{INT}_i &= e_{int} + h_{int}(\text{PAT}_i) + \varepsilon_{int,i} \\
\text{DSout}_i &= e_{int} + e_{ds} + h_{ds}(\text{PAT}_i) + \varepsilon_{ds,i}
\end{align*}$$ (11)

The first equation states that patient $i$’s observed intermediate outcome comprises of a component $e_{int}$ that represents provider effort on intermediate outcome, a component $h_{int}(\text{PAT}_i)$ (a function of $\text{PAT}_i$) due to patient $i$’s existing health condition, and a residual error $\varepsilon_{int,i}$. The second equation provides an analogous decomposition for the observed downstream outcome. Therefore $\text{INT}_i - h_{int}(\text{PAT}_i) = e_{int} + \varepsilon_{int,i}$ is patient $i$’s risk-adjusted intermediate outcome, and $\text{DSout}_i - h_{ds}(\text{PAT}_i) = e_{int} + e_{ds} + \varepsilon_{ds,i}$ is patient $i$’s risk-adjusted downstream outcome.
A short caveat is in place here. Equation (11) admits an interesting physiological interpretation: Suppose the effect of intermediate effort on downstream outcomes is not direct but indirect through its effect on intermediate outcomes, which in turn affect downstream outcomes. That is, intermediate effort improves intermediate outcomes (dialysis dosage and anemia) and these improvements then lead to improvements in downstream outcomes. In this case, the intermediate and downstream outcomes would admit the following relationship:

$$DS_{out_i} = e_{ds_i} + h'_{ds}(PAT_i) + INT_{out_i} + \varepsilon'_{ds,i}$$  \hspace{1cm} (12)$$

Equation (12) then leads into the equation for $DS_{out_i}$ in (11) if we substitute $e_{int} + h_{int}(PAT_i) + \varepsilon_{int,i}$ into the place of $INT_{out_i}$ in the right hand side of (12).

We will also assume that the provider expends the same effort level $(e_{int}, e_{ds})$ in treating each patient, the cost of effort is independent of each patient characteristics and denoted $g(e) = c^T e + \frac{1}{2}e^T Q e$, the error terms $\{\varepsilon_i\}^N_{i=1} \equiv \{(\varepsilon_{int,i}, \varepsilon_{ds,i})\}^N_{i=1}$ are independent draws from (2), and the agent’s utility remains exponential. The attentive reader will notice that unlike the conceptual model in section 2, the model here assumes a cost of effort with both a linear and quadratic term. The linear term is included because it was shown in the estimation part of this paper to provide a better fit to the data than a pure quadratic model. Furthermore, the cost function does not include a fixed cost but this does not affect the analysis: Specifically, assume the fixed cost is $g_0$, then this cancels out from the incentive compatibility constraint (5). In the individual rationality constraint (6), the reservation utility $U_0$ is the agent’s utility under the current system. This is equal to the agent’s utility in the current system assuming $g_0 = 0$, denoted $\hat{U}_0$, minus the unknown fixed cost $g_0$. If we substitute for $U_0 = \hat{U}_0 - g_0$ in the right hand side of (6), then the $g_0$s cancel from both sides of the equation leading into an individual rationality constraint without a fixed cost and with reservation utility of $\hat{U}_0$.

Medicare’s current payment system is

$$s_{current} = \pi_{current} \sum^N_{i=1} DS_{out_i}$$  \hspace{1cm} (13)$$

while a linear payment system analogous to (3) is:

\[ s = N\pi_0 + \pi_{int} \sum_{i=1}^{N} (INT_{out_i} - h_{int}(PAT_i)) + \pi_{ds} \sum_{i=1}^{N} (DS_{out_i} - h_{ds}(PAT_i)). \]  \( (14) \)

Analysis analogous to that used for the model in section 2 shows that the principal-agent model for this formulation can be analyzed by solving the single patient problem and closed form expressions can be derived for the optimal payment system. In the process, we can also obtain expressions for the agent’s optimal effort under the current payment system, denoted \( e^c = (e^c_{int}, e^c_{ds}) \). Then, the agent’s expected utility under that system is the reservation utility:

\[ \hat{U}_0 = \pi_{current}(e^c_{int} + e^c_{ds}) - g(e) - \frac{r}{2}\sigma^2_{ds}\pi_{current}. \]  \( (15) \)

**Implementation.** To implement the proposed system (14) we will need to estimate the following parameters: the functions \( f \) defined in (10) and \( h \) defined in (11), the provider’s current effort levels \( (e^c_{int}, e^c_{ds}) \), error variance \( \Sigma \) (2), cost function \( g \), and reservation certainty equivalent (15). Finally to calculate the optimal payment rates \( (\pi_0, \pi_{int}, \pi_{ds}) \) we will also need, in addition to these quantities, the provider’s degree of risk aversion \( r \) and the reward \( \nu \) that Medicare receives for each year of hospital-free life (downstream outcome). We will show in section 4 how to estimate all these parameters, except \( r \) and \( \nu \), from patient-level data. We will propose and justify baseline parameters for \( r \) and \( \nu \) in section 6, where we will also perform extensive sensitivity analysis.

4. **Statistical Estimation**

We will now extend the model from section 3 into one with multiple providers in order to estimate its main parameters from patient data. We first describe the nature of the data that will be used for parameter estimation, next we extend (11) to the multiple provider setting and also derive a first order condition for the agent’s optimal effort under the current system. Finally, we use these equations to derive seven estimating equations (moment conditions) that can be used to estimate all model parameters.

**Data format for multiple providers.** We now assume that we have data from \( N \) patients treated in one of \( M \) dialysis providers. The subscript index \( i \in 1, \ldots, N \) tracks the \( i^{th} \) patient and...
the superscript index \( j \in 1, \ldots, M \) denotes the \( j \)th dialysis provider. Let \( N^j \) denote the number of patients treated by provider \( j \). For each patient \( i \), we have data covering a complete year of treatment and we observe the following for that year: a) Downstream outcomes \( DSout_i \); b) The percentage of treatments during the observation period with dosage exceeding the 65% threshold, \( DOSAGE_i \); c) The percentage of treatments with hematocrit in the optimal range 33% to 36% \( ANEMIA_i \); d) The intermediate outcomes \( INTout_i = f(DOSAGE_i, ANEMIA_i) \) (the exact form of the function will be specified later); e) The co-variate vector \( PAT_i \); f) A \( K \)-dimensional choice vector \( C_i \) where its \( j \)th entry, \( C^j_i \), is 1 if patient \( i \) is treated in facility \( j \) and zero otherwise.

For future references, it will also be necessary to define the \( K + M + 2 \)-dimensional vector \( Z_i = (PAT_i, C_i, DOSAGE_i, ANEMIA_i) \). In addition, for each provider \( j \) we observe its revenue rate for each unit of downstream outcome under the current system, denoted \( \pi^j_{\text{current}} \), and the revenue rate varies across providers.

**Generalization of (11) and First-Order Condition.** With the notation for the multiple providers in place, we can now generalize (11) as follows

\[
\begin{align*}
INTout_i &= \sum_{j=1}^{M} C^j_i e^j_{\text{int}} + h_{\text{int}}(PAT_i) + \varepsilon_{\text{int},i} \\
DSout_i &= \sum_{j=1}^{M} C^j_i e^j_{\text{ds}} + \sum_{j=1}^{M} C^j_i e^j_{\text{int}} + h_{\text{ds}}(PAT_i) + \varepsilon_{\text{ds},i}
\end{align*}
\] (16)

where \( (e^j_{\text{int}}, e^j_{\text{ds}}) \) is the unobserved effort level of the \( j \)th provider. Also, under the current payment system, each agent’s unobserved effort level will maximize the agent’s certainty equivalent.

Straightforward algebra shows that this leads to the following equations for provider \( j \):

\[
\begin{align*}
\pi^j_{\text{current}} - c_1 - Q_{11} e^j_{\text{int}} - Q_{12} e^j_{\text{ds}} &= 0 \\
\pi^j_{\text{current}} - c_2 - Q_{12} e^j_{\text{int}} - Q_{22} e^j_{\text{ds}} &= 0
\end{align*}
\] (17)

Equations (16) and (17) are the basic equations that describe the relationship between the unobserved agent effort and outcomes, and between the effort and the agent’s cost and reward. These equations will be used to derive the seven estimation equations (also referred to as moment conditions) to be presented next, but first it is necessary to provide additional specification for two components of (16): the definition of the functions \( f(DOSAGE_i, ANEMIA_i) \) and \( h_{\text{int}}(PAT_i) \). For
the function that defines \( INT_{out_i} \) we will assume a flexible form given by the tensor product of piecewise linear functions:

\[
INT_{out_i}(\beta) \equiv f_\beta(DOSAGE_i, \text{ANEMIA}_i) = \sum_{j,k=1}^{4} \beta_{jk}(DOSAGE - D_j)_+(ANEMIA - A_k)_+ \quad (18)
\]

where \( x_+ = \max(0, x) \) and \( \{D_j\}_{j=1,...,4} \) and \( \{A_k\}_{k=1,...,4} \) are the quartiles of \( DOSAGE \) and \( \text{ANEMIA} \) respectively. \( \beta \) is the regression coefficient to be estimated with maximum likelihood estimate \( \hat{\beta} \).

Modeling \( h \) using tensor products is not practical because of the high dimensionality of the vector \( \text{PAT}_i \). An alternative is to use a generalized additive model for \( h \) (Hastie & Tibshirani (7)) in which the vector of patient variables \( \text{PAT}_i \) is transformed to a set of basis vectors \( \text{PAT}_i^{\text{GAM}} \) such that \( h(\text{PAT}_i) \) can be expressed as \( \theta^T \text{PAT}_i^{\text{GAM}} \). The transformation \( \text{PAT}_i^{\text{GAM}} \) is predetermined and the vector \( \theta \) is to be estimated. Incorporating the functional forms \( f \) and \( h \) into (16) gives

\[
INT_{out_i}(\beta) = \sum_{j=1}^{M} C_j^i e_{int}^j + \theta_1^T \text{PAT}_i^{\text{GAM}} + \varepsilon_{\text{int},i} \quad \text{DSout}_i = \sum_{j=1}^{M} C_j^i e_{ds}^j + \sum_{j=1}^{M} C_j^i e_{ds}^j + \theta_0^T \text{PAT}_i^{\text{GAM}} + \varepsilon_{ds,i}. \quad (19)
\]

**Estimation.** Starting with equations (17) and (19) we will now show how to obtain the maximum likelihood estimates for the unknown model parameters: \( \beta, \theta_0, \theta_1, e_{int}^j, e_{ds}^j, c, Q \), and \( \Sigma \); the estimates will be denoted \( \hat{\beta}, \hat{\theta}, \hat{e}, \hat{c}, \hat{Q} \) and \( \hat{\Sigma} \), respectively. We will use the Empirical Likelihood method (EL; Owen (14)) which proceeds as follows: First, we derive seven moment functions from equations (17) and (19). These moment functions will be denoted \( m_k(i|\beta, \theta, e, c, Q, \Sigma) \); \( k = 1, ..., 7 \). These are essentially functions whose expectation under the maximum likelihood estimator probability model will be zero. The probability model is nonparametric: the observations come from an unknown probability distribution \( \{p_1, \ldots, p_N\} \) on the \( N \) observations/patients. Then, the estimates are obtained by finding the parameters \( (\beta, \theta, e, c, Q, \Sigma) \) and the probability distribution \( \{p_1, \ldots, p_N\} \) that maximizes the log-likelihood

\[
\sum_{i=1}^{N} \log(p_i), \quad (20)
\]
subject to the constraint that the expectation of all moment functions is zero (we refer to these constraints as the moment conditions):

$$
E \sum_{i=1}^{N} p_i m_k(i|\beta, \theta, e, c, Q, \Sigma) = 0; k = 1, ..., 7
$$

(21)

where $p_i : i = 1, ..., N$ is a probability distribution:

$$
\sum_{i=1}^{n} p_i = 1 \text{ and } p_i \geq 0; i = 1, ..., N.
$$

(22)

The estimates for the probability distribution are essentially nuisance parameters, while the estimators $\hat{\beta}, \hat{\theta}, \hat{e}, \hat{c}, \hat{Q}$ and $\hat{\Sigma}$ have the same optimality properties as parametric maximum likelihood estimators: they are consistent and asymptotically normal. Also, this method provides a likelihood ratio test that can be used for hypothesis testing. Furthermore, compared to alternative estimation methods that are commonly used in econometrics, such as the Generalized Method of Moments, EL not only has the same first-order asymptotic properties (which means the methods are equivalent for very large sample sizes) but it has better higher order properties (which mean that it is better for smaller samples sizes). The reader is referred to Newey and Smith (13) for further details.

The moment functions. The seven moment functions are as follows (the moment functions for the third to seventh moment conditions have $M$ dimensions, one for each provider; thus, functions for provider $j$ are denoted $m_3^j(i|\beta, \theta, e, c, Q, \Sigma) - m_7^j(i|\beta, \theta, e, c, Q, \Sigma)$, respectively):

$$
m_1(i|\beta, \theta, e, c, Q, \Sigma) = (INT_{out}(i) - \frac{\theta_i^T PAT_i^{GAM}}{\theta_0^T PAT_i^{GAM}}) Z_i,
$$

$$
m_2(i|\beta, \theta, e, c, Q, \Sigma) = (DS_{out} - \sum_{j=1}^{M} C_j e_{int,j} - \sum_{j=1}^{M} C_j e_{ds,j} - \frac{\theta_0^T PAT_i^{GAM}}{\theta_0^T PAT_i^{GAM}}) Z_i,
$$

$$
m_3^j(i|\beta, \theta, e, c, Q, \Sigma) = C_j^i \left[ (INT_{out}(i) - \sum_{j=1}^{M} C_j e_{int,j} - \sum_{j=1}^{M} C_j e_{ds,j} - \frac{\theta_0^T PAT_i^{GAM}}{\theta_0^T PAT_i^{GAM}})^2 - (\sigma_j^i)^2 \right],
$$

(23)

$$
m_4^j(i|\beta, \theta, e, c, Q, \Sigma) = C_j^i \left[ (DS_{out} - \sum_{j=1}^{M} C_j e_{int,j} - \sum_{j=1}^{M} C_j e_{ds,j} - \frac{\theta_0^T PAT_i^{GAM}}{\theta_0^T PAT_i^{GAM}})^2 - (\sigma_j^i)^2 \right] \times \left( INT_{out}(i) - \sum_{j=1}^{M} C_j e_{int,j} - \sum_{j=1}^{M} C_j e_{ds,j} - \frac{\theta_0^T PAT_i^{GAM}}{\theta_0^T PAT_i^{GAM}} \right),
$$

$$
m_5^j(i|\beta, \theta, e, c, Q, \Sigma) = C_j^i \left[ \pi_{current} - c_1 - Q_{11} e_{int,1} - Q_{12} e_{ds,1} \right],
$$

$$
m_6^j(i|\beta, \theta, e, c, Q, \Sigma) = C_j^i \left[ \pi_{current} - c_2 - Q_{12} e_{int,2} - Q_{22} e_{ds,2} \right].
$$

Several remarks regarding the detailed derivation and interpretation of these functions are now in place. The first two functions are derived from (16). They specify the functional forms of $f$ and $h$, and estimate the providers’ current efforts. In these equations, $INT_{out}(i) - \sum_{j=1}^{M} C_j e_{int,j}$
\( \theta_i^T PAT_i^{GAM} \) is the residual from the equation defining the intermediate outcomes and \( DS_{out_i} = \sum_{j=1}^M C_i^j e_{int}^j - \sum_{j=1}^M C_i^j e_{ds}^j - \theta_i^T PAT_i^{GAM} \) is the residual from the equation for the downstream outcomes. \( Z_i \) is the \( K \)-dimensional vector of co-variates and thus, both moment functions have \( K \)-dimensions. These two functions collectively generate a total of \( 2K \) equations which state that the residuals should be uncorrelated with the co-variates vector \( Z_i \). The next three sets of equations (derived from (16)) specify the covariance structure \( \Sigma \). Each of the these three equations has \( M \) dimensions, one dimension for each provider. This generates a total of \( 3M \) moment equations. These functions state that the variance matrix for the residuals gives the covariance matrix \( \Sigma \), and they assume a different covariance matrix for each provider. Finally, the last two equations (derived from the agent’s first-order conditions) specify the cost-of-effort function for the dialysis providers. They have \( M \) dimensions, one for each provider, and they represent the first order condition for the intermediate effort \( (m_i^6(i | \beta, \theta, e, c, Q, \Sigma)) \) and downstream effort \( (m_i^4(i | \beta, \theta, e, c, Q, \Sigma)) \). These last two functions imply that the expectation for the first order conditions of each provider is taken only on the patients treated by that provider. Collectively, the seven moment functions (23) reflect the following four assumptions: A) The function defining the intermediate outcomes and the risk adjustment is the same for all providers; B) The covariance matrix is different for each provider; C) The first order conditions are solved separately by each provider; D) All providers face the same "average" cost of effort function (i.e. \( c \) and \( Q \) are the same for all \( M \) providers) but different rewards.

A final remark concerning the solution of (20)-(22). This optimization problem has a total of \( 2K + 5M \) equality constraints and \( N + K + 5M + 21 \) unknowns (\( N \) for the probability distribution, \( 2M \) efforts, \( 3M \) variance parameters, \( K \) for risk-adjustment, 5 from the cost of effort function and 16 from the tensor product). Because \( K \) is expected to be much smaller than \( N + 21 \), the unknowns will far exceed the equality constraints and there will be multiple feasible solutions. The optimal solution is the probability distribution that satisfies all constraints and generates the largest log-likelihood.
Implementation of the Optimal Contract (14). The estimation method provides all parameters needed to design the optimal linear payment system described in (14) with the exception of the agent’s risk aversion and principal’s reward per unit of downstream outcomes (more about these two quantities will be said in section 6). Beyond that, the effort estimates obtained from our methods can provide the performance measures used in the payment system. Specifically, the payment system (14) is equivalent (within a constant) to a system that rewards provider $j$ based on

$$s^j = N^j \left[ \pi_0^j + \pi_{\text{int}}^j \hat{e}_j^\text{int} + \pi_{\text{ds}}^j (\hat{e}_j^\text{int} + \hat{e}_j^\text{ds}) \right].$$

(24)

The estimated intermediate effort, $\hat{e}_j^\text{int}$, for provider $j$ is equal to the average over all patients treated by that provider of the difference between observed and expected intermediate outcomes $(\text{INTout}_i - h_{\text{int}}(\text{PAT}_i))$ utilized in (14). Similarly, $\hat{e}_j^\text{int} + \hat{e}_j^\text{ds}$ is the average over all patients treated in provider $j$ of $(\text{DSout}_i - h_{\text{ds}}(\text{PAT}_i))$. Furthermore, both $\hat{e}_j^\text{int}$ and $\hat{e}_j^\text{int} + \hat{e}_j^\text{ds}$ are normally distributed with respective means $e_j^\text{int}$ and $e_j^\text{int} + e_j^\text{ds}$, and with an expression for the variance given in the online companion. Therefore, the estimation method provides all relevant information for the optimal design of the payment system (24), and the outcomes used in (24) follow the assumption of normality assumed in the stylized model. To put it differently, the stylized models described in Sections 2 and Sections 3 provide a reasonable representation of a principal-agent model based on (24): the observed performance measures $\hat{e}_j^\text{int}$ and $\hat{e}_j^\text{ds}$ have a normal distribution and are centered at the unobserved agent’s effort. The most stringent assumptions that will not be verified empirically here are that of exponential utility and that the cost of effort is the same for all providers. Relaxing these assumptions is possible, but left for future research.

One last caveat is in order. In our Medicare example, provider effort levels must obey a constraint not introduced in the conceptual models presented in Sections 2 and 3: the agent’s feasible effort pair should be restricted in the range for which we have adequate data to estimate the agent’s cost of effort function. Such a range is provided by the 95% confidence ellipsoid of estimated efforts, and this can be obtained from our estimation method. The optimal payment system is then derived using constrained numerical optimization over the ellipsoid.
5. Data Sources

Data were obtained from the United States Renal Data System (USRDS (19)), a database that includes all patients covered by Medicare’s ESRD program. Our analysis included \( N = 8,160 \) ESRD patients in New England who received in-center hemodialysis at any time in 2003 and for whom Medicare was the primary payer. These patients received dialysis from \( M = 150 \) dialysis facilities. For each patient, the data provided a set of basic demographic information as well as medical information such as the disease causing ESRD and other co-existing diseases that could be used for risk-adjustment. Hospital days were also provided, including hospital costs, and dialysis provider payments. For each dialysis treatment provided to each patient, the database included information on the dialysis dose (Urea Reduction Ratio) and anemia management (Hematocrit Level). The hematocrit and dialysis dose variables were used to define the variables \( DOSAGE_i \) and \( ANEMIA_i \).

Downstream outcomes, \( DSout_i \), were calculated as the fraction of hospital-free days in 2003. For patients who started dialysis before Jan 1, the calculation of \( DSout_i \) was straightforward and was given by the total number of hospital-free days divided by 365; for example, a patient who was alive and on dialysis for the whole year with one month in the hospital had \( DSout_i = 11/12 \), a patient who died on Dec 1 but was on dialysis for all eleven preceding months and had one month in the hospital had \( DSout_i = 10/12 \) (subtract two months from the year, one for death and one for hospital-stays). The calculation was more subtle for patients who started dialysis after Jan 1st. In their case, we started with the fraction of the year during which the patient was alive and then multiplied that by the fraction of the year on dialysis when the patient was hospital free. This method attempted to take into account all available information for each patient and correct for the lack of a complete one year of observation after the start of dialysis. Because any method of handling these incomplete observations could introduce bias, we repeated our analysis using two alternative approaches: a) include only patients who were already on dialysis by Jan 1; b) include all patients but define \( DSout_i \) as the fraction of the time interval between the initiation of dialysis (if it occurred after Jan. 1) and the end of year during which the patient was not hospitalized and
alive (this is a modification of the baseline method which ignores the fact that patient was alive prior to the initiation of dialysis).

Summary statistics for all the variables used in our estimation method are presented in the online companion, Table EC.2. The summary statistics provide estimates for some of the parameters that cannot be estimated using the moment conditions (23). Specifically, Medicare’s cost for hospitalization is seen to be $30,600 per patient year at risk and Medicare’s payment for one full year of uninterrupted dialysis services is $27,900. This implies that the principal’s reward for a full hospital-free year is \( \nu = 30,600 \) and that the average parameters for the current payment system are \( \pi_0^c = 0, \pi_{ds}^c = 27,900 \) and \( \pi_{int}^c = 0 \); in fact, the parameters for the payment system varied by provider. These parameters will be used in the next section to design a new payment system.

6. Results

Intermediate Score Function, Intermediate and Downstream Effort, and Cost of Effort.

Figures 1 and 2 present three-dimension plots of estimated intermediate outcome score function \( INT_{out}(\hat{\beta}; DOSAGE, ANEMIA) \) (18) and estimated cost function, respectively. Figure 3 is a scatterplot of the pair of estimated provider effort \( (\hat{e}_{int}^i, \hat{e}_{ds}^i) \) for all providers in the data set; all effort levels are normalized to be non-negative.

![Figure 1](image)

**Figure 1** Three dimensional plot of the intermediate outcome score \( INT_{out}(\hat{\beta}; DOSAGE, ANEMIA) \) as a function of \( DOSAGE \) and \( ANEMIA \).
$g(e) = \begin{pmatrix} 2.28 \\ 2.51 \end{pmatrix}^T e + \frac{1}{2} e^T \begin{pmatrix} 1.10 & 0.41 \\ 0.41 & 0.33 \end{pmatrix} e$

(Ajusted $R^2 = 0.0343$)

Figure 2  Three dimensional plot of the estimated cost of effort $g(e_{int}, e_{ds})$ as a function of intermediate effort, $e_{int}$, and downstream effort $e_{ds}$.

Figure 3  Scatterplot of the estimated pairs of intermediate and downstream effort for each provider. The ellipsoid gives a 95% confidence interval.

Several observations are in place here:

1) The intermediate outcome score function is jointly non decreasing and concave in the two intermediate quality measures. This implies diminishing returns for the effort a provider puts into raising intermediate quality measures. More significantly, $ANEMIA$ values greater than 8% have little effect on the score function, implying that increases in ANEMIA beyond 8% were not associated with a change in the Downstream outcomes and hence will not be rewarded by the
optimal payment system.

2) Intermediate and downstream efforts are substitutes: increasing one, makes the cost of increasing the second one higher, and this makes increases in the second one less desirable for the agent (assuming linear rewards).

3) There is wide variability across providers in effort levels. The mean effort vector was \((\hat{e}_{\text{int}} = 0.264, \hat{e}_{\text{ds}} = 0.466)\) with the 95% confidence band for the intermediate effort ranging between 0.13 and 0.38 while the corresponding band for the downstream effort was between 0.21 and 0.63: a 0.42 hospital free years difference between the lowest end and the upper end of the 95% confidence band.

4) The goodness of fit for the cost of effort function can be improved significantly. With an adjusted \(R^2\) of 0.034 it appears that a more flexible functional forms and incorporation of provider characteristics (such as chain affiliation, size, etc) could lead into an improved model.

5) Improving downstream outcomes through a change in downstream effort was, on average, almost as expensive as improving it through a change in the intermediate effort. Specifically, an increase of 0.10 in either the intermediate effort or the downstream effort would translate into an expected increase in the downstream outcome of hospital-free years by 0.10. Using the estimated cost function, we calculated that the corresponding incremental cost of effort (starting from the cost of the average effort level which was $17,837) were $2,799 for the intermediate effort and $2,779 for the downstream effort. The savings from Medicare from this increase in the either intermediate or downstream effort would be $30,600 \times 0.10 = $3,060. Since Medicare’s potential savings exceed the incremental cost of effort, it implies that there are opportunities to reward better downstream outcomes in a way that would cover the provider’s incremental cost without increasing overall Medicare expenditures.

Covariance Matrix. The estimated average (over all providers) variance for the error term in the intermediate outcomes was \(\hat{\sigma}_{\text{int}}^2 = 0.0204 \pm 3.08 \times 10^{-4}\) years and for the downstream outcomes was \(\hat{\sigma}_{\text{ds}}^2 = 0.0931 \pm 7.91 \times 10^{-4}\) years. The correlation between intermediate and downstream outcomes was 0.425. Since the variance in the intermediate outcomes noise was less than that in the
downstream outcomes, incorporating intermediate outcomes in the payment system is expected to be beneficial for the principal.

**Hypothesis testing.** The preceding discussion, viewed in light of the conceptual models in sections 2 and 3, suggests that a payment system that better rewards intermediate and downstream outcomes may improve patient outcomes without increasing Medicare’s costs. Some of the critical assumptions driving this conclusion are a) that intermediate outcomes are associated with patient downstream outcomes and b) they alone cannot explain variations in the downstream outcomes (there is a downstream effort component that affects downstream outcomes beyond the intermediate effort). The statistical analysis we performed here also generates two hypothesis tests that can be used to assess the validity of these assumptions. To evaluate the first assumption, we used the empirical likelihood ratio test (see online companion) to test the null hypothesis that \( INTout_i = 0 \) i.e. \( H_0 : \beta = 0 \) versus the alternative \( H_A : \beta \neq 0 \). We found significant evidence \((p\text{-value} < 0.001)\) to reject the null, providing empirical support for the proposal to reward providers based on their performance on these measures. For the second assumption, we tested whether \( c_{ds}^j = 0 \) for \( j = 1, \ldots, M \). We found that the intermediate outcomes alone did not explain all the variations in downstream outcome \((p\text{-value} 0.0042)\).

**Sensitivity Analysis.** We performed parameter estimation using the two alternative methods for handling patients with incomplete observations described in Section 5 (excluding patients with incomplete observations, and using alternative normalizations for patients with incomplete observations). We calculated correlation coefficients between the estimates derived using the three methods (baseline and two alternatives) for the intermediate effort, downstream effort, covariance estimate, parameters in the intermediate score function, in the cost of effort function, and in the coefficients for the co-variates in the risk-adjustment. Correlation coefficients were high. The lowest one was 0.85 between the estimates for the provider’s downstream effort using the baseline approach and using the method that dropped incomplete observations. A total of eighteen pairwise correlation coefficients were calculated: seven were between 0.85 and 0.90 (three of these came from the downstream effort and two from the intermediate effort); two were between 0.90 and 0.95,
and the remaining were above 0.95. This suggests that the estimates for the model parameters are, in general, insensitive to the method used to handle incomplete observations, while the effort estimates for individual providers are more sensitive. This is not surprising since some providers may have a limited number of observations and hence their effort estimates are expected to be more volatile.

**Optimal payment system for Medicare.** We will now use the parameter estimates derived above in order to obtain the optimal pay-for-performance system for Medicare’s ESRD program that incorporates both intermediate and downstream outcomes. This will be contrasted to the following alternatives: (a) Medicare’s current system; b) A system based only on downstream outcome ((24) but with $\pi_{int} = 0$); c) A system based only on intermediate outcome ((24) but with $\pi_{ds} = 0$). We will present results for a fictitious dialysis provider with the average covariance structure and cost function identified above. In addition, our analysis will assume that Medicare’s reward is $\nu = $30,600 per hospital-free year. For the agent’s risk aversion, we will consider the following values: $r = 0.00002, 0.0001, 0.0003$ and $0.0004$ per dollar respectively; $r = 0.00002$ will be the baseline. The interpretation of these coefficients is easier in the form of the certainty equivalent for a gamble in which the risk-averse provider has a 50% chance of winning $1,000 and a 50% chance of losing $1,000. The certainty equivalent values corresponding to $r = 0.00002, 0.0001,$ $0.0003$ and $0.0004$ are $-$10, $-$50, $-$100, $-$150 and $-$200, respectively. The results for the four systems are presented in Table 1.

Several remarks are in place here:

1. The effort pair induced by the current system, $(e_{int} = 0.295, e_{ds} = 0.457)$, is very close to the actual average provider effort $(e_{int} = 0.264, e_{ds} = 0.466)$ estimated from the USRDS data. This confirms the models considered here are internally consistent.

2. The payment system that rewards downstream outcomes alone achieves almost the same performance as Medicare’s current payment system. This suggests that through years of experimentation Medicare has already optimized its system as much as it can be optimized using downstream outcomes alone.
Table 1 Description and Summary of the Performance of Alternative Payment Systems.

<table>
<thead>
<tr>
<th>r = 0.00002/$</th>
<th>Current reimbursement system</th>
<th>Reward intermediate outcomes only</th>
<th>Reward downstream outcomes only</th>
<th>Reward both outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ν = $30,600/year</td>
<td>$0 per year</td>
<td>$7,250 per year</td>
<td>$0 per year</td>
<td>$0 per year</td>
</tr>
</tbody>
</table>

Payment for each unit of $e_{int}$: $\pi_{int} + \pi_{ds}$

| Payment for each unit of $e_{ds}$: $\pi_{ds}$ | $27,900 per year | $0 per year | $28,000 per year | $28,100 per year |

Induced effort $(e_{int}, e_{ds})$

| (0.295, 0.457) | (0.300, 0.271) | (0.291, 0.491) | (0.222, 0.604) |

Expected intermediate outcome

67% sufft. dosage 71% sufft. dosage 65% sufft. dosage 65% sufft. dosage

| Expected downstream outcome per patient | 0 (baseline) | −0.181 yr | +0.0300 yr | +0.0740 yr |

Cost to provider per patient per year

| $0 (baseline) | $5,000 | +$838 | +$2,090 |

Provider revenue per patient per year

| $0 (baseline) | $5,570 | +$920 | +$2,140 |

Provider net earnings per patient per year

| $0 (baseline) | $573 | +$77 | +$55 |

Medicare savings per patient per year

| $0 (baseline) | +$30 | +$2.80 | +$123 |

3. The optimal system pays $200 more for each unit of $e_{ds}$ than the current system ($28,100 vs. $27,900), but $200 less for each unit of $e_{int}$ ($27,700 vs. $27,900). This induces a decrease of 0.295 − 0.222 = 0.073 years in provider’s intermediate effort and an increase of 0.604 − 0.457 = 0.147 years in downstream effort, generating a net increase of 0.074 hospital-free years per patient, or 27 days per year. An annual savings of $123 per patient for Medicare was realized and the provider’s net earnings per patient were higher than under the current system (even though the provider’s risk was higher). In this new system, the total expected cost for the provider increased from $19,573 under the current system to $21,663, a net increase of $2,080; while the provider’s expected revenue increased by $2,140. At the same time, Medicare’s expected cost from the downstream outcomes decreased by 0.073 x $30,600 = $2,260, resulting into a net saving for Medicare of $123.

This demonstrates that there is an opportunity to improve overall patient outcomes by increasing the total expected payment for downstream outcomes while slightly decreasing the total expected payment for intermediate outcomes. The net increase in Medicare’s payment to the provider will
be compensated by a net decrease in Medicare’s hospitalization expenditures.

4. While the difference in payment rates between the optimal system and the current system appears to be small, the total change in Medicare’s annual payment for each patient was a very significant $2,140. This increase was primarily driven by the increase of 0.074 in $DSOUT_i$ which translated into more treatments and more payments for the dialysis provider. That is, the small change in the payment rate was magnified because the payment rate was multiplied by $DSOUT_i$.

5. Two factors drive observations (3) are (4): First, the net cost of increasing provider effort in the region around current effort levels is less than the net savings for Medicare from improved downstream outcomes. Therefore, Medicare can increase total provider payments to reward better downstream outcomes, without increasing its total cost (Payments + Cost of Hospitalization). Second, in the estimated cost of effort function, the marginal cost of the downstream effort is less than the marginal cost of intermediate effort. Thus, additional savings are possible by differentiating between the two types of effort, and rewarding improvements in downstream effort more than corresponding improvements in the intermediate one.

7. Intermediate outcomes alone were not sufficient to generate downstream outcomes equivalent to the current system, despite their lower variance. The optimal system needed to rely on both intermediate and downstream outcomes.

**Sensitivity analysis.** We investigated the performance of the optimal payment system when the provider became increasingly risk-averse and when Medicare’s reward for the downstream outcome $\nu$ increased. We found that as the agent’s coefficient of risk-aversion increased, the principal supplemented the agent’s payment by a modest fixed payment (the highest was $71 per patient) in order to meet the agent’s reservation utility. The more interesting observations were made when Medicare’s reward varied. First, for values greater than $40,000, the induced second-best effort pair was beyond the 95% confidence band. Second, a $5,000 increase in Medicare’s reward per hospital-free year translated into a 0.073 increase in hospital-free years, but an additional $5,000 increase beyond that generated a much smaller increment of 0.011 hospital-free years. These translate into $68,993 per hospital-free year gained for the first $5,000 and $454,545 for the next $5,000. Since
an increase of Medicare’s reward beyond its current level would require an increase in Medicare’s budget, this analysis implies that a modest increase in Medicare’s ESRD budget ($5,000 per hospital free year) would be cost effective if used to strengthen the provider’s incentives to improve downstream outcomes.

7. Concluding Remarks

We have developed an empirical framework to estimate the parameters of a multi-task principal-agent model. The framework was applied to patient data from Medicare’s ESRD program and was used to address the following questions: a) How can intermediate process-compliance measures be integrated into a single intermediate performance score to be used in an optimal payment system? b) What is the agent’s cost of intermediate and downstream effort and reservation utility? c) Is it possible to improve the existing payment system? The methodology developed is general and can be applied to data from other systems, beyond Medicare’s ESRD program. These could include large networks of service providers in financial services (e.g. bank branches), automobile dealer networks etc. The main requirement is that the data should include observations from multiple agents with different payment rates or payment systems in order to estimate the relative differences between the agents’ efforts.

Our results for Medicare’s program suggest that an optimal payment system can extend the hospital-free days of dialysis patients in New England by almost one month per year while reducing Medicare’s expenditures by $123 per patient per year. Extrapolating this finding to all Medicare beneficiaries receiving dialysis treatment from freestanding providers in the U.S. (265,958 patients) this would translate into a gain of 19,700 hospital-free life years along with savings of $32.7M. This is a significant potential improvement in patient outcomes without a net increase in Medicare expenditures.

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, for the pay-for-performance system developed here to work, the patient outcomes used to determine compensation must be properly risk-adjusted by patient health factors that are beyond
the control of the dialysis provider. For otherwise there would be an incentive for the provider to ‘cream-skim’ healthier patients in pursuit of better outcomes, creating a perverse situation where sicker patients have reduced access to healthcare. Although we used a large number of known patient health conditions for risk adjustment in estimating the provider effort $e$, the presence of unobserved but influential patient health factors could cause endogeneity problems that would lead to inconsistent estimates. McClellan, McNeil & Newhouse (11) introduced the use of patient travel distances to healthcare facilities as instrumental variables (Bowden & Turkington (1)) to correct for endogeneity problems in healthcare provider quality assessment. We could use the same instruments here: in the first two moment conditions in (23) the co-variate vector of instruments $Z_i$ can be replaced with a vector of differential distances (defined as the vector of the difference in distance between each patient’s residence and each facility, minus the distance to the closest facility). This can significantly reduce selection bias from the estimation of agent’s effort but at the expense of more computations and less efficient estimates (instrumental variable estimators have higher variance that our original estimators). The latter means that the resulting incentive system derived from the instrumental variable estimates will involve higher risk for the agent (and thus will be more expensive for the principal) than the incentive system without the instrumental variable adjustment. We ultimately decided against the use of instrumental variables for the following reason: Results in a separate paper (Lee & Zenios (8)), where we used differential distance and applied instrumental variable estimation to identify dialysis facility-specific factors that correlate with provider quality, have shown only weak evidence for the existence of endogeneity in dialysis patient outcomes for 2003 (Hausman’s specification test p-value = 0.10). Given the weak evidence we judged that any reduction in bias from using instrumental variable techniques would be unlikely to counterbalance the loss of efficiency, and thus was not justified.

Second, our analysis assumed that the agent’s objective was to maximize profit alone. This could be viewed as a very cynical assumption: after all, health care providers are governed by the principles of medical ethics which dictate that patient interests should come ahead of the provider’s own self-interest. In future work, we intend to extend the framework to empirically examine whether
a model in which the agent’s utility includes outcomes beyond the agent’s profit, such as pure patient outcomes, provides a better fit to the data.

Third, our analysis assumed that all providers in the data had the same cost function. In reality, we would expect variations across providers based on provider attributes such as size, type of ownership (for-profit, non-profit, chain-affiliated versus independent) and others. Investigation of these effects is also left for future research.

Four, the cost function in our analysis assumed that the cost of delivering dialysis treatments to each patient does not vary with the number of treatments provided. A more appropriate model would be to assume that \( g(e) = g_0(e) + g_1(e)DSout_t \). That is, there is a fixed cost per patient that does not vary with the number of treatments and a variable cost that reflects supplies for each treatment. We have examined two options for such a function: in one we assumed that the variable cost was constant and did not vary with the agent’s effort (and the results were very similar to what was reported in section (6)); in the second we assumed that the variable component depended on the agent’s effort but the resulting estimated function was not monotonic. This implies that additional work needs to be undertaken to validate our assumptions for the cost function. Similar limitation applies to our assumption that the payment rate in the current system is independent of effort. In reality, providers can increase their payment rate by more aggressive administration of injectable drugs such as EPO. A model capturing that effect can be developed but the first order conditions for the agent’s best response to the current system will be more complex (and possibly analytically intractable). Therefore, refinement of the model for the current payment system is also left for future research.

Five, we have been unable to derive moment conditions to estimate the provider’s degree of risk aversion. Although our sensitivity analysis showed that the payment rates associated with the optimal reimbursement system is invariant over a large range of risk aversion values, we intend to examine in the future different methods to identify estimating equations involving \( r \). We conjecture that an empirical analysis of a dynamic models such as (Plambeck & Zenios (15)) or static models with a non-constant coefficient of risk aversion may provide such an opportunity. Also, in estimating
the provider’s cost-of-effort function from the first order condition (17), we assumed a quadratic functional form for convenience. Ideally, we would want a flexible function that is non-decreasing in both arguments. Spline-based techniques could provide the desired flexibility but finding the best-fit model while ensuring that it is increasing in both effort arguments is a very difficult task; see for example (Wood (17)) and (Leitenstorfer & Tutz (9)) for the simpler univariate case.

The results from the optimal payment system presented in section 6 suggest that relatively minor tweaking of the current reimbursement system can lead into significant improvements in outcomes. This is a very strong conclusion, therefore it is worth revisiting the key assumptions driving it: The optimal reimbursement system was derived from the estimated cost function, which in turn was developed using the assumption of profit-maximizing agents. Therefore, to gain more confidence about our conclusion we will need to establish the correctness of the estimated cost function and the appropriateness of the pure profit maximization assumption. For the latter, we will fit in future research models where the agent’s utility will depend both on its profit and on patient outcomes. For the former, we can cross-validate the estimated cost function to cost reports provided to Medicare by all dialysis facilities. In these cost reports, the facilities break down their costs by their sources and provide estimates for the cost per treatment. We can use these data to first augment our moment condition with one additional condition: that the average cost per patient under the estimated cost function matches data in the cost report. And then we will test whether the cost functions derived using these expanded moment conditions can prospectively predict future costs and provider effort as reported by dialysis facilities to Medicare.

In order for the results from this analysis to influence the practice of dialysis and the design of dialysis incentive systems, it will be necessary to link our abstract estimates for intermediate and downstream effort to actual actions the providers take. There are two approaches that can be taken: One, we can use the cost reports for each provider (containing information on staffing levels, drug acquisition costs, capital expenditures etc) to examine whether differences in estimated efforts are correlated with specific items in the cost reports. Second, providers with higher than average (lower than average) estimated effort can be surveyed to identify any activities they perform that
can explain the higher (lower) than average performance. Once such activities are identified (i.e. hypotheses are generated), a study can be designed to empirically test such hypotheses.

In summary, we have taken a small first step towards developing an empirical methodology to design evidence-based optimal incentive systems. Our method estimates the parameters of a multi-task principal-agent using observations from multiple agents with different payment rates. An application using data from ESRD patients treated in multiple dialysis facilities has demonstrated the feasibility of this method, while highlighting several important questions that can form the basis for future research. With the abundance of data made possible by modern IT systems, we believe that the problem of developing data-driven methods to design incentives systems can be a very fruitful area of research for Operations Management and Management Science scholars. The applications are numerous and extend beyond the domain of health services.

Acknowledgments

The authors would like to thank Laurence Baker, Margaret Brandeau, Michael Harrison, Brian Lee, Chris Lee, Art Owen and Alan Sorensen for their helpful suggestions. Donald Lee was supported by a Croucher Foundation Scholarship.

References


This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.
Illustrative Applications of the Empirical Methodology

Applications of this empirical methodology are numerous and the extend beyond the ESRD example in the manuscript. Three additional applications are outlined now:

- Automobile sales. The auto manufacturer (principal) benefits when a customer chooses a car from the same manufacturer to replace his or her earlier purchase (downstream outcome). The manufacturer relies on its franchised dealerships (agents) to provide its customers with a quality experience before and after the purchase of a vehicle. The end-goal is maximizing customer retention. While the likelihood of a customer’s repeat business is certainly influenced by his or her experiences at a dealership, other factors also play a role. Hence it is hard to discern a dealership’s service effort from repeat businesses alone. Also, measuring repeat purchases requires long periods of observations, and may be challenging due to the lack of data. However the manufacturer can use surveys to gauge a customer’s experience with a dealership along some dimensions, and also determine whether it complied with certain customer relations protocols. The question is then to design a pay-for-performance system that synthesizes information from these surveys of customer satisfaction into a single performance measure that is aligned with customer retention. Rewarding the dealer based on that measure would then induce actions that maximize the principal’s downstream outcome (customer retention).

- Gaming. In order to maximize customer satisfaction and minimize customer defections to competitors, casinos offers complimentary such as free flight tickets, rooms and cash rebates to valued players. The casino is the principal and the player the agent, the downstream outcome is the casino’s profits. Typically, the value of the complimentary issued is a percentage of the casino’s winnings from the player. However, the player can use betting strategies (effort) that increase the volatility of the game, and of the casino’s profit, to a point where such loss-rebate programs can cause losses for the casino (Lucas, Kilby & Santos (10)). Casino management can observe part of the player’s strategy, such as wagers and number of games played (intermediate outcomes). Rebating players based on both actual losses and intermediate outcomes can reduce volatility and ensure that the casino is profitable, without compromising customer satisfaction.
- Healthcare. The ESRD example is a special case of the payer-provider interaction that is prevalent in health care delivery. A healthcare provider (agent) treats a patient and is reimbursed by the principal (the patient or his/her insurance carrier). High quality treatment reduces future complications (negative downstream outcome), but the incidence of complications also depends on the patient’s existing health-related risk factors, behavior, and other factors. The principal also observes the agent’s adherence to certain clinical processes and best practices, for example the dosage and frequency of treatment or frequency of monitoring tests. These are the intermediate outcomes and the question is to identify the optimal mix of intermediate and downstream outcomes to include in a performance-based payment system.

**Background on Empirical Likelihood Estimation**

The Empirical Likelihood method due to (Owen (14)) provides a very natural way for solving regression problems via the use of moment conditions, and has certain advantages over Generalized Method of Moments. We show how to apply this method to our specific problem at hand. The core of the idea is that the \( N \) observations (patients) \( \{Z_i, DS_{out_i}\}_{i=1}^{N} \) come from an unknown distribution \( P \). Suppose that the \( i^{th} \) datapoint occurs with probability \( p_i \), where \( \sum_{i=1}^{N} p_i = 1 \). The empirical log-likelihood takes on the form

\[
I_{EL}(p) = \sum_{i=1}^{N} \log p_i
\]  

(EC.1)

and attains its maximum value at \( p_i = 1/N \). Now suppose that we impose a model on the observed data, which further constrains the distribution by requiring that \( p \in S \subset \mathbb{R}^N \). In analogy to the idea of a likelihood ratio for parametric families, the nonparametric empirical likelihood ratio can be defined in a similar manner:

\[
ELR = \max_p \left\{ \sum_{i=1}^{N} \log(p_i N) \left| p \in S, p_i \geq 0, \sum_{i=1}^{N} p_i = 1 \right. \right\}
\]  

(EC.2)

In section 4 we obtained a set of estimating equations \( E[m_k(i|\beta, \theta, e, c, Q, \Sigma)] = 0 \) for \( k = 1, ..., 7 \) that must be satisfied by the regression coefficients \( (\beta, \Sigma, e, g, \theta) \). Given specific values for
$(\beta, \Sigma, e, g, \theta)$, the empirical moment conditions $\sum_{i=1}^{N} p_i m_k(i|\beta, \theta, e, c, Q, \Sigma) = 0$ imposes a constraint on $p$, so the empirical likelihood ratio for our model is

$$ELR(\beta, \theta, e, c, Q, \Sigma) = \max_{p} \left\{ \sum_{i=1}^{N} \log p_i \left| \sum_{i=1}^{N} p_i m_k(i|\beta, \theta, e, c, Q, \Sigma) = 0, p_i \geq 0, \sum_{i=1}^{N} p_i = 1 \right. \right\}$$  \hspace{1cm} (EC.3)

Since $(\beta, \theta)$ are nuisance parameters that do not show up in the solution to the principal-agent mode, we can define the reduced likelihood ratio to obtain

$$ELR(e, c, Q, \Sigma) = \max_{p, \beta, \theta} \left\{ \sum_{i=1}^{N} \log p_i \left| \sum_{i=1}^{N} p_i m(i|\beta, \Sigma, e, \theta) = 0, p_i \geq 0, \sum_{i=1}^{N} p_i = 1 \right. \right\}$$  \hspace{1cm} (EC.4)

and the maximum empirical likelihood estimates (MELE) $\hat{e}, \hat{c}, \hat{Q}, \hat{\Sigma}$ are the values that maximize the reduced likelihood ratio (EC.4). Under mild regularity conditions (Qin & Lawless (16)) shows that $(\hat{e}, \hat{c}, \hat{Q}, \hat{\Sigma})$ is consistent and asymptotically normal with limiting variance

$$\lim_{N \to \infty} \frac{1}{N} \left[ \mathbb{E} [\nabla_e c, Q, \Sigma m]^T \mathbb{E} [mm^T] \mathbb{E} [\nabla_e c, Q, \Sigma m] \right]^{-1}$$  \hspace{1cm} (EC.5)

In addition to having optimal efficiency and bias properties, the empirical likelihood ratio statistic is also appealing because it enjoys all the nice statistical properties of a parametric likelihood ratio. For example suppose we wish to test the null hypothesis that $e = 0$ vs. $e \neq 0$. The likelihood ratio test based on the asymptotic result

$$-2ELR(0, \hat{c}, \hat{Q}, \hat{\Sigma}) \xrightarrow{d} \chi^2_{d.f.(e)}$$  \hspace{1cm} (EC.6)

carries over from the parametric case and provides an easy way to test hypotheses. When instrumental variables are used, common diagnostic tools like the test of overidentifying restrictions and Hausman-Wu specification tests can also be constructed from the empirical likelihood ratio in a very natural manner.
Supporting Tables

<table>
<thead>
<tr>
<th>Helper variable $J$</th>
<th>(a) intermediate &amp; downstream</th>
<th>(b) downstream only</th>
<th>(c) intermediate only</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J$</td>
<td>$\begin{pmatrix} 1 &amp; 1 \ 0 &amp; 1 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 0 &amp; 1 \ 0 &amp; 1 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 1 &amp; 0 \ 0 &amp; 0 \end{pmatrix}$</td>
</tr>
<tr>
<td>Helper variable $K$</td>
<td>$J^T Q^{-1} J - r \Sigma$</td>
<td>$J^T Q^{-1} J - r \sigma_{ds}^2$</td>
<td>$J^T Q^{-1} J - r \sigma_{int}^2$</td>
</tr>
<tr>
<td>Helper variable $L$</td>
<td>$J^T Q^{-1} J + r \Sigma$</td>
<td>$J^T Q^{-1} J + r \sigma_{ds}^2$</td>
<td>$J^T Q^{-1} J + r \sigma_{int}^2$</td>
</tr>
<tr>
<td>$\pi^*_0$</td>
<td>$\begin{pmatrix} \pi^<em>_0 \ \pi^</em>_{ds} \end{pmatrix}$</td>
<td>$\begin{pmatrix} \pi^<em>_0 \ \pi^</em>_{ds} \end{pmatrix}$</td>
<td>$\begin{pmatrix} \pi^<em>_0 \ \pi^</em>_{ds} \end{pmatrix}$</td>
</tr>
<tr>
<td>$(e^<em>_int, e^</em>_ds)$</td>
<td>$Q^{-1} J \begin{pmatrix} \pi^<em>_0 \ \pi^</em>_{ds} \end{pmatrix}$</td>
<td>$Q^{-1} J \begin{pmatrix} \pi^<em>_0 \ \pi^</em>_{ds} \end{pmatrix}$</td>
<td>$Q^{-1} J \begin{pmatrix} \pi^<em>_0 \ \pi^</em>_{ds} \end{pmatrix}$</td>
</tr>
</tbody>
</table>
## Table EC.2  Summary of data used in the analysis: patient attributes, clinical and financial outcomes.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description (if applicable)</th>
<th>Data Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Age on January 1st 2003.</td>
<td>Quartiles: [60.5, 71.4, 78.8] years</td>
</tr>
<tr>
<td>Gender</td>
<td>55.7% male</td>
<td></td>
</tr>
<tr>
<td>Time since dialysis initiation</td>
<td>Time spent on dialysis therapy as of January 1st 2003.</td>
<td>Quartiles: [0.00, 1.08, 2.80] years</td>
</tr>
<tr>
<td>Employment</td>
<td>11.7% of patients are employed.</td>
<td></td>
</tr>
<tr>
<td>Insurance coverage</td>
<td>Coverage at the onset of kidney disease. Options include None, Medicare/Medicaid, Veteran’s Admin. and Other (options are not mutually exclusive).</td>
<td>None: 4.41%, Medicare/Medicaid: 78.0%, VA: 1.66%, Other: 55.6%</td>
</tr>
<tr>
<td>Body mass index</td>
<td>Measure of body fat based on height and weight</td>
<td>Quartiles: [22.5, 25.9, 30.6]</td>
</tr>
<tr>
<td>Serum albumin level</td>
<td>Measure of nutrition levels. Low albumin levels indicate malnourishment (≤3.5g/dl).</td>
<td>Quartiles: [2.80, 3.30, 3.60] g/dl</td>
</tr>
<tr>
<td>Serum creatinine level</td>
<td>Indicator of kidney function.</td>
<td>Quartiles: [5.1, 6.8, 8.8] mg/dl</td>
</tr>
<tr>
<td>Blood urea nitrogen level</td>
<td>Indicator of kidney function.</td>
<td>Quartiles: [71, 92, 116] mg/dl</td>
</tr>
<tr>
<td>Hemoglobin level</td>
<td>Indicator of kidney function.</td>
<td>Quartiles: [8.9, 9.9, 11.0] g/dl</td>
</tr>
<tr>
<td>Glomerular filtration rate</td>
<td>Indicator of kidney function.</td>
<td>Quartiles: [5.92, 7.95, 10.7] ml/min</td>
</tr>
<tr>
<td>Ability to move</td>
<td>96.1% of patients</td>
<td></td>
</tr>
<tr>
<td>Tobacco use</td>
<td>4.91% of patients</td>
<td></td>
</tr>
<tr>
<td>Alcohol use</td>
<td>1.15% of patients</td>
<td></td>
</tr>
<tr>
<td>Drug dependence</td>
<td>0.760% of patients</td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>7.19% of patients</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>22.8% of patients</td>
<td></td>
</tr>
<tr>
<td>History of hypertension</td>
<td>78.7% of patients</td>
<td></td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>13.5% of patients</td>
<td></td>
</tr>
<tr>
<td>Pericarditis</td>
<td>1.04% of patients</td>
<td></td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td>10.0% of patients</td>
<td></td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
<td>19.3% of patients</td>
<td></td>
</tr>
<tr>
<td>Downstream outcome</td>
<td>% of the year 2003 that a patient was alive and hospital-free</td>
<td>Quartiles: [60.5, 71.4, 78.8] years</td>
</tr>
<tr>
<td>Dialysis adequacy</td>
<td>% of treatments with a Urea Reduction Ratio of ≥65%</td>
<td>Quartiles: [56%, 78%, 91%]</td>
</tr>
<tr>
<td>Sufficient anemia management</td>
<td>% of treatments with hematocrit levels between 33% and 36%</td>
<td>Quartiles: [8.0%, 32%, 50%]</td>
</tr>
<tr>
<td>Medicare payments to providers for dialysis</td>
<td>Payment rate for one year of dialysis if patient is never hospitalized</td>
<td>Quartiles: [$25,700, $27,900, $30,200] per patient per year</td>
</tr>
<tr>
<td>Medicare payments for hospital expenditures</td>
<td>Total hospitalization expenditures per patient in 2003</td>
<td>Quartiles: [$10,000, $20,800, $39,600] per patient. Mean=$30,600/patient</td>
</tr>
</tbody>
</table>
## Table EC.3 Reference index for commonly used notations

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ANEMIA_i$</td>
<td>Fraction of time that patient $i$‘s hematocrit level stayed between 33% to 36%</td>
</tr>
<tr>
<td>$DOSAGE_i$</td>
<td>Fraction of time that patient $i$‘s blood urea reduction exceeded 65%</td>
</tr>
<tr>
<td>$DSout_i$</td>
<td>Downstream outcome of patient $i$. Proportion of time that patient was not hospitalized</td>
</tr>
<tr>
<td>$e_{ds}$</td>
<td>Component of provider $j$’s efforts that influence patient downstream outcome</td>
</tr>
<tr>
<td>$e_{int}$</td>
<td>Component of provider $j$’s efforts that influence patient intermediate outcome</td>
</tr>
<tr>
<td>$g(e)$</td>
<td>Agent’s cost function for supplying effort level $e = (e_{int}, e_{ds})$. In practice we approximate this with its $2^{nd}$ order Taylor approximation $g(e) = g_0 + c^T e + \frac{1}{2} e^T Q e$</td>
</tr>
<tr>
<td>$i$</td>
<td>Index for labeling dialysis patients, $i \in {1, \ldots, N}$</td>
</tr>
<tr>
<td>$INTout_i$</td>
<td>Intermediate outcome score of patient $i$ (see (10), (18))</td>
</tr>
<tr>
<td>$j$</td>
<td>Index for labeling dialysis providers, $j \in {1, \ldots, M}$</td>
</tr>
<tr>
<td>$M$</td>
<td>Number of dialysis providers in study</td>
</tr>
<tr>
<td>$N, N_j$</td>
<td>Number of dialysis patients in study and number of patients treated by provider $j$</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Reward that Medicare/principal receives for each unit of downstream outcome $DSout$ (1 year of no hospitalizations for a patient) ‘produced’ by provider/agent</td>
</tr>
<tr>
<td>$PAT_i, PAT_{iGAM}$</td>
<td>Set of observable patient attributes that are used to risk-adjust for patient $i$‘s intermediate and downstream outcomes</td>
</tr>
<tr>
<td>$\pi^{current}_j$</td>
<td>Revenue rate for each unit of downstream outcome under the current system</td>
</tr>
<tr>
<td>$Q$</td>
<td>Quadratic term of $2^{nd}$ order Taylor approximation to cost function $g(e)$</td>
</tr>
<tr>
<td>$r$</td>
<td>A measure of agent’s of risk aversion (see provider’s utility $U$)</td>
</tr>
<tr>
<td>$s^{current,s}$</td>
<td>Medicare’s current and proposed payment system contract (see (13), (14))</td>
</tr>
<tr>
<td>$\Sigma, \sigma_{int}^2, \sigma_{int,ds}^2, \sigma_{ds}^2$</td>
<td>See (2)</td>
</tr>
<tr>
<td>$U$</td>
<td>Provider’s utility function $U(x) = -e^{-rx}$</td>
</tr>
<tr>
<td>$U_0$</td>
<td>Reservation certainty equivalent of agent (15)</td>
</tr>
<tr>
<td>$Z_i$</td>
<td>The vector (instrument) of independent regression variables of patient $i$: $(DOSAGE, ANEMIA, C, PAT_{iGAM})$</td>
</tr>
</tbody>
</table>