2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Optimal Contract Design for End-of-Life Care Payments

Muyan Jiang, Ying Chen, Xin Chen, Javad Lavaei, and Anil Aswani

2024 IEEE 63rd Conference on Decision and Control (CDC) | 979-8-3503-1633-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/CDC56724.2024.10886745

Abstract— **A large fraction of total healthcare expenditure occurs due to end-of-life (EOL) care, which means it is important to study the problem of more carefully incentivizing necessary versus unnecessary EOL care because this has the potential to reduce overall healthcare spending. This paper introduces a principal-agent model that integrates a mixed payment system of fee-for-service and pay-for-performance in order to analyze whether it is possible to better align healthcare provider incentives with patient outcomes and cost-efficiency in EOL care. The primary contributions are to derive optimal contracts for EOL care payments using a principal-agent framework under three separate models for the healthcare provider, where each model considers a different level of risk tolerance for the provider. We derive these optimal contracts by converting the underlying principal-agent models from a bilevel optimization problem into a single-level optimization problem that can be analytically solved. Our results are demonstrated using a simulation where an optimal contract is used to price intracranial pressure monitoring for traumatic brain injuries.**

I. INTRODUCTION

End-of-life (EOL) care is a large part of the total spending on healthcare. For instance, in the United States, a significant portion of health expenditures occurs in the final year of life, comprising more than 13% of Medicare spending [1], with sharply increasing costs in the last few days of life, especially for hospitalized patients. Despite the enormous spending on EOL care, studies suggest these expenditures do not necessarily improve patient outcomes [2]–[4]. Typically, EOL decisions involve discussions with stakeholders, including family or the patients themselves. Early EOL conversations are associated with less aggressive and more cost-effective care, reducing expenditures by up to 95% [5], and palliative care can lower hospitalization costs and the likelihood of readmission [6]. Given the significant disconnect between the current level of spending on EOL care and patient outcomes, it is important to study further how to balance these issues better.

Various payment models, including capitation, salary, fee-for-service (FFS), pay-for-performance (P4P), and combinations of these, have been proposed for healthcare. Changes in payment methods can significantly impact the cost-effectiveness of clinical decisions [7]. In EOL care, there is a push to integrate P4P incentives to improve palliative care quality [8]. Meanwhile, empirical evidence indicates a preference for the FFS model for terminally ill patients, particularly over other Medicare plans [9].

*This material is based upon work supported by the National Science Foundation under Grant DGE-2125913 and Grant CMMI-1847666.

MJ, YC, XC, JL, and AA are with the Department of Industrial Engineering & Operations Research, University of California, Berkeley, CA 94720, USA {muyan_jiang, ying-chen, xin.chen2024, lavaei,aaswani}@berkeley.edu

A. Related Work

Various studies have examined contract design for medical topics like incentivizing medical data sharing [10]. Research has also investigated how medical choices affect healthcare costs in chronic disease patients, focusing on their interaction with principal-agent dynamics [11]. Common-agency problems in the US healthcare system have been analyzed for their effects on contracting and care coordination [12]. Additionally, the design of financial incentives in integrated care, especially in bundled care models, has been explored [13]. Overall, advanced Medicare payment models are associated with improved patient quality of life, cost savings, and provider satisfaction [14]. However, to the best of our knowledge, optimal payment pricing for EOL care has not been studied.

B. Payment Models for EOL Care

The main challenge in designing efficient payment models for EOL care is the information asymmetry between the payer (e.g., health insurance or government) and the healthcare provider. The provider knows more about the patient's condition, while the payer cannot directly assess whether the EOL care provided was necessary. The payer only sees the efforts (procedures and costs) and the patient outcomes (e.g., survived or died within 30 days). Consequently, payers often use fee-for-service models, compensating providers based on the costs of performed procedures.

The setup of the above-described situation motivates the core idea of our paper: Suppose we design a contract where the payment amount to a provider depends upon both the amount of effort and the outcome of the patient. Would such a contract better incentivize providers to only provide necessary EOL care? The idea is that the contract could be designed such that it gives the highest payments when the provider exerts effort and the patient has a good outcome, whereas lower payments are provided when the patient has a poor outcome. By modulating payments for effort by the outcome, our proposed contract combines elements of fee-for-service and pay-for-performance payment models.

C. Contributions and Outline

Our paper makes three contributions. First, to the best of our knowledge, the idea of modulating EOL care payments based on a binary outcome measure is novel. The underlying principal-agent models we propose are also new. Second, we derive the optimal contracts for these models. Although these models are structured as bilevel optimization problems, with the payer at the upper level designing a contract to maximize utility and the provider at the lower level maximizing their utility under the contract, we simplify this to a single-level optimization by incorporating constraints that represent healthcare providers' utility. This approach allows us to analytically solve the optimization problem and derive the optimal contract. Third, we develop an algorithm to estimate the parameters of our optimal contracts, demonstrating its effectiveness with real data.

Section II defines the utility functions for the payer and provider and formulates three principal-agent models, each reflecting different levels of provider risk tolerance. In Section III, we solve the optimization problems for these models and derive the optimal contracts, providing practical insights. Finally, Section V demonstrates our framework with a numerical simulation set in the context of intracranial pressure monitoring for traumatic brain injuries, introducing a novel algorithm for estimating model parameters.

II. MODEL FORMULATIONS

In the healthcare system, providers deliver EOL care to patients and receive compensation from payers. Patients are categorized as $S = 1$ for favorable/good responders and $S = 0$ for unfavorable/bad responders. Good responders are those whose profiles suggest greater benefits from intensive treatment. Based on patients' conditions, healthcare providers decide on the expenditure level E, choosing between high-cost intensive interventions ($E = 1$) or lower-cost palliative care ($E = 0$). Patient outcomes are denoted as $Q = 1$ for survival and $Q = 0$ for mortality which depends on both the responder status and the treatment given. The reimbursement $P = p_{ij}$ depends on both the outcome $Q = i$ and the expenditure level $E = j$ for $i, j \in \{0, 1\}$. Thus, providers must carefully weigh both the quality and cost aspects of end-of-life care.

A. Payer's Utility Function

We assume a Bernoulli model for the patient's outcome Q depending on responder status $S = s$ and expenditure level $E = j$, $Q \sim \text{Bern}(\pi_{sj})$ for $s, j \in 0, 1$. The probability of a good responder status is also assumed as another independent Bernoulli model $S \sim \text{Bern}(\gamma)$ for $\gamma \in (0, 1)$. These simplifications are made to focus on gaining initial insights into the relationship between expenditure and outcomes. While more complex models could be considered, our approach allows for clearer interpretation and analytical tractability.

Assumption 1. $\pi_{01} \geq \pi_{00}, \pi_{11} \geq \pi_{10}, \pi_{10} \geq \pi_{00}, \pi_{11} \geq \pi_{01}$

Remark 1. *This indicates that higher expenditure generally results in a better survival rate. Moreover, a good responder yields a higher survival rate than a bad responder at the same expenditure level. Additionally, under the Bernoulli assumption,* $0 < \pi_{01},\ \pi_{00},\ \pi_{11},\ \pi_{10},\ \gamma < 1$.

The payer's utility is a weighted sum of the expected survival rate of the patients and the expected payment:

$$u_{\text{payer}} = \mathbb{E}(Q) - \phi \cdot \mathbb{E}(P), \tag{1}$$

where $\mathbb{E}(Q) = (1-\gamma) \cdot (1-\pi_{00}) + \gamma \cdot \pi_{11}$ and $\mathbb{E}(\cdot)$ denotes expectation. Here, that constant parameter $\phi > 0$ is a weight.

B. Provider's Utility Function

The provider incurs a disutility F for high expenditure. Without loss of generality, we normalize the units of the principal-agent models by assuming that $F = 1$. Considering the payments and costs, the provider's utility is

$$u_{\text{provider}} = \mathbb{E}(P|E) - \mathbb{1}(E = 1) \cdot F \tag{2}$$

where $\mathbb{1}(\cdot)$ is an indicator function that equals one when the condition inside is true and zero otherwise.

C. Principal-Agent Model Formulations

We construct principal-agent models where the payer is the principal, and the provider is the agent. We constrain our model to ensure the expenditure level aligns with the patient type; that is, the healthcare provider will exert intensive care on good responders and palliative care otherwise.

1) Free Payment Model: The first model is to maximize the payer's utility while allowing the payer to *fine* the provider, if desired at optimality, for bad outcomes:

$$\begin{aligned} \max_{p_{ij}} \quad & \mathbb{E}(Q) - \phi \cdot \mathbb{E}(P) \\ \text{s.t.} \quad & \mathbb{E}(P) \geq 0 \\ & \mathbb{E}(P|S=1, E=1) - F \geq \mathbb{E}(P|S=1, E=0) \\ & \mathbb{E}(P|S=0, E=0) \geq \mathbb{E}(P|S=0, E=1) - F \end{aligned} \tag{3}$$

The first constraint in (3) ensures that the healthcare provider does not incur any loss (in expectation) in this scheme.

2) Non-Negative Payment Model: The second model does not allow the payer to fine the provider where we require all payments to be non-negative:

$$\begin{aligned} \max_{p_{ij}} \quad & \mathbb{E}(Q) - \phi \cdot \mathbb{E}(P) \\ \text{s.t.} \quad & \mathbb{E}(P|S=1, E=1) - F \geq \mathbb{E}(P|S=1, E=0) \\ & \mathbb{E}(P|S=0, E=0) \geq \mathbb{E}(P|S=0, E=1) - F \\ & p_{00}, p_{01}, p_{10}, p_{11} \geq 0 \end{aligned} \tag{4}$$

We also examine the optimal solution under scenarios where a false diagnosis of responder status could occur.

3) Risk-Averse Agent Model: A risk-averse provider has diminishing marginal utility for payments, meaning their utility function is concave. This implies they prefer a certain payment over a risky one with the same expected value.

$$\begin{aligned} \max_{p_{ij}} \quad & \mathbb{E}(Q) - \phi \cdot \mathbb{E}(P) \\ \text{s.t.} \quad & \mathbb{E}(g(P)|S=1, E=1) - F \geq \mathbb{E}(g(P)|S=1, E=0) \\ & \mathbb{E}(g(P)|S=0, E=0) \geq \mathbb{E}(g(P)|S=0, E=1) - F \\ & p_{00}, p_{01}, p_{10}, p_{11} \geq 0 \end{aligned} \tag{5}$$

where $g(\cdot)$ is a bijective concave function that is positively valued whenever its argument is a positive value.

III. OPTIMAL CONTRACTS

Next, we design optimal contracts for the models described above by solving the corresponding optimization problems.

A. Free Payment Model

For every feasible incentive design, $\mathbb{E}(Q)$ is a constant with respect to the outcome probabilities. In this situation, we want the optimal solution with the smallest expected reimbursement. The preferred optimal solution can be reduced to solving the following system of equations:

$$\begin{cases} \mathbb{E}(P) = 0 \\ \mathbb{E}(P|S=1,E=1) - 1 \geq \mathbb{E}(P|S=1,E=0) \\ \mathbb{E}(P|S=0,E=0) \geq \mathbb{E}(P|S=0,E=1) - 1 \end{cases}$$

Denote

$$\vec{c_0} = [(1-\gamma)(1-\pi_{00}), \gamma(1-\pi_{11}), (1-\gamma)\pi_{00}, \gamma\pi_{11}]^T,$$
$$\vec{c_1} = [\pi_{10} - 1, 1-\pi_{11}, -\pi_{10}, \pi_{11}]^T,$$
$$\vec{c_2} = [1-\pi_{00}, \pi_{01} - 1, \pi_{00}, -\pi_{01}]^T,$$
$$b_0 = 0, b_1 = 1, b_2 = -1$$

With the definition for $\vec{P} = \begin{bmatrix} p_{00} & p_{01} & p_{10} & p_{11} \end{bmatrix}^T$, the problem is equivalent to solving:

$$\begin{cases} \vec{c_0} \cdot \vec{P} = b_0 \\ \vec{c_1} \cdot \vec{P} \geq b_1 \\ \vec{c_2} \cdot \vec{P} \geq b_2 \end{cases} \tag{6}$$

Proposition 1. *The binding system of* (6) *admits a solution if and only if* $(1-\gamma)\pi_{01} + \gamma\pi_{11} < 1$.

Proof. For the system

$$\begin{bmatrix} \vec{c_0}^T \\ \vec{c_1}^T \\ \vec{c_2}^T \end{bmatrix} \vec{P} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

One can do row reduced echelon on the coefficient matrix $\begin{bmatrix} \vec{c_0}^T \\ \vec{c_1}^T \\ \vec{c_2}^T \end{bmatrix}$ and arrive at:

$$\begin{pmatrix} 1 & 0 & 0 & -\frac{(\pi_{01}-\pi_{11})((\gamma-1)\pi_{00}-\gamma\pi_{10})}{(\pi_{00}-\pi_{10})((\gamma-1)\pi_{01}-\gamma\pi_{11}+1)} \\ 0 & 1 & 0 & \frac{1}{(\gamma-1)\pi_{01}-\gamma\pi_{11}+1} - 1 \\ 0 & 0 & 1 & -\frac{(\pi_{01}-\pi_{11})((\gamma-1)\pi_{00}-\gamma\pi_{10}+1)}{(\pi_{00}-\pi_{10})((\gamma-1)\pi_{01}-\gamma\pi_{11}+1)} \end{pmatrix}$$

which shows that the matrix has full rank under the condition $(1-\gamma)\pi_{01} + \gamma\pi_{11} < 1$. □

Theorem 1. *One optimal solution of* (6) *is*

$$\begin{aligned} p_{00} \to & [-p_{11}(\pi_{11} - \pi_{01})s_0 + \gamma(\pi_{00}\pi_{01} - 2\pi_{00}\pi_{11} \\ & + \pi_{00} + \pi_{10}\pi_{11} - \pi_{10}) + \pi_{00}(\pi_{11} - \pi_{01})] \\ & /[(\pi_{10} - \pi_{00})(1 - s_1)], \\ p_{01} \to & (-p_{11}s_1 - \gamma + 1)/(1 - s_1), \\ p_{10} \to & [p_{11}(\pi_{11} - \pi_{01})(1 - s_0) + \gamma(\pi_{00}\pi_{01} - 2\pi_{00}\pi_{11} \\ & + \pi_{00} - \pi_{01} + \pi_{10}\pi_{11} - \pi_{10} + \pi_{11}) - (1 - \pi_{00}) \\ & \cdot (\pi_{11} - \pi_{01})]/[(\pi_{10} - \pi_{00})(1 - s_1)], \end{aligned}$$

where $s_0 = (1-\gamma)\pi_{00} + \gamma\pi_{10} < 1$ *and* $s_1 = (1-\gamma)\pi_{01} + \gamma\pi_{11} < 1$ *are expected survival rate for high and low expenditure respectively. Notice they increase as* γ *increases.*

Proof. After obtaining the reduced row echelon form in the previous proposition, one can derive the linear solution subspace with p_{11} as a free variable. □

Remark 2. *Regarding* p_{00}, p_{01}, p_{10} *as a function of* p_{11} *we have*

$$\frac{\partial p_{00}}{\partial p_{11}} = -\frac{(\pi_{11} - \pi_{01})s_0}{(\pi_{10} - \pi_{00})(1 - s_1)} < 0,$$
$$\frac{\partial p_{01}}{\partial p_{11}} = -\frac{s_1}{1 - s_1} < 0,$$
$$\frac{\partial p_{10}}{\partial p_{11}} = \frac{(\pi_{11} - \pi_{01})(1 - s_0)}{(\pi_{10} - \pi_{00})(1 - s_1)} > 0$$

which means that as we increase the welfare of high expenditure spending with desirable outcomes, to maintain optimality, we need to increase that of low expenditure of survival outcome and decrease others.

Proposition 1 shows that an optimal contract ensures zero expected payment for the payer while upholding the constraints that ensure low expenditures for bad responders and high expenditures for good responders. The zero expected payment is achievable in this model because the payer is able to fine the provider when outcomes are poor.

B. Non-Negative Payment Model

The optimization problem in (4) can be reformulated to

$$\begin{aligned} \min_{\vec{P}, \vec{S}} \quad & \begin{bmatrix} c_0^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \vec{P}^T & \vec{V}^T \end{bmatrix}^T - b_0 \\ \text{s.t.} \quad & \begin{bmatrix} c_1^T & -1 & 0 \\ c_2^T & 0 & -1 \end{bmatrix} \begin{bmatrix} \vec{P}^T & \vec{V}^T \end{bmatrix}^T = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\ & \begin{bmatrix} \vec{P}^T & \vec{V}^T \end{bmatrix} \geq 0 \end{aligned} \tag{7}$$

where $\vec{V} = \begin{bmatrix} v_1 & v_2 \end{bmatrix}^T$ are two slack variables.

Assumption 2. $\pi_{01}\pi_{10} \neq \pi_{00}\pi_{11}$

Remark 3. *The assumption* $\frac{\pi_{10}}{\pi_{00}} \neq \frac{\pi_{11}}{\pi_{01}}$ *means that the transitional benefit in survival rate from good responder to bad responder is different with different expenditure levels.*

Theorem 2. *Under Assumptions 1 and 2, the optimal objective value of* (7) *is* $m^* = \gamma$*, with solution of the following form:*

$$p_{00} \to 0, p_{01} \to t, p_{10} \to 0, p_{11} \to \frac{1}{\pi_{11}} - \frac{1-\pi_{11}}{\pi_{11}}t,$$

where $0 \leq t \leq 1$.

Proof. By adding slack variables, the optimal solution of model (7) must satisfy four active linearly independent constraints. After checking all basic points, and under Assumptions 1 and 2, there are only two solutions both primal feasible and dual feasible, which are

$$p_{00} = 0, p_{01} = 1, p_{10} = 0, p_{11} = 1, v_1 = 0, v_2 = 0,$$
$$p_{00} = 0, p_{01} = 0, p_{10} = 0, p_{11} = \frac{1}{\pi_{11}}, v_1 = 0, v_2 = 1 - \frac{\pi_{01}}{\pi_{11}}$$

The optimal solution lies on the line segment of these two basic points, with optimal value $m^* = \gamma$. □

The optimal contract (2) for the non-negative payment model is characterized by a parameter t, and the quality of the contract is equivalent for any $t \in [0,1]$. In the special case $t = 0$, the provider receives zero payment if there is low expenditure or if there is high expenditure but a bad outcome. The provider only receives a payment when there is a high expenditure and a good outcome.

In practical scenarios, the assumption of perfect classification of responders is not realistic. To address this, we introduce two parameters: $w_0 := \Pr(S = 0|TS = 1)$, representing the false negative rate for responder classification, and $w_1 := \Pr(S = 1|TS = 0)$, denoting the false positive rate, where TS is the true responder status class. These parameters encapsulate the inaccuracies inherent in the medical classification process. Importantly, this information remains concealed from the healthcare provider, who makes decisions based only on the observed responder class and chooses the corresponding expenditure level. Consequently, this modification affects the objective value by changing the survival distribution, impacting the payer's utility.

Specifically, the new model is the same as (7) except the objective becomes $c_0^{wT}\vec{P} - b_0$ where

$$\begin{aligned} c_0^{wT} = [&(1-w_0)(1-\gamma)(1-\pi_{00}) + w_1\gamma(1-\pi_{10}),\\ &w_1(1-\gamma)(1-\pi_{01}) + (1-w_0)\gamma(1-\pi_{11}),\\ &(1-w_1)(1-\gamma)\pi_{00} + w_0\gamma\pi_{10},\\ &w_1(1-\gamma)\pi_{01} + (1-w_0)\gamma\pi_{11}]^T. \end{aligned}$$

The coefficient is formulated based on the unchanged constraint to incentivize healthcare providers to treat observed favorable responders.

Proposition 2. *In* (7)*, replacing c_0 with c_0^w, under Assumptions 1 and 2, we obtain the optimal solution at*

$$p_{00} = 0, p_{01} = 0, p_{10} = 0, p_{11} = \frac{1}{\pi_{11}}, v_1 = 0, v_2 = 1 - \frac{\pi_{01}}{\pi_{11}}$$

with optimal objective value $m_w{}^* = \gamma\left(1 - \frac{\pi_{01}w_1}{\pi_1} - w_0\right) + \frac{\pi_{01}w_1}{\pi_1}$.

Proof. Similar to Theorem 2, checking all basic points will reveal the only optima in this case. □

Interestingly, the optimal contract for this modified model where patients' responder status may be misclassified by the provider is the same as the contract corresponding to $t = 0$ for the original form of this model with exact patient statuses.

Remark 4. *From Theorem 2 and Proposition 2, direct algebra reveals that*

$$m^* < m_w^* \Leftrightarrow \frac{\pi_{01}w_1}{\pi_{11}w_0 + \pi_{01}w_1} > \gamma$$

C. Risk-Averse Agent Model

Considering the model in (5), we can define g^{-1} since it is a bijective function. Relabeling $W = g(P)$, and noting the survival rate is again a constant under the constraints, we can reformulate the problem as follows:

$$\begin{aligned} \min_{W} \quad & c_0^T g^{-1}(W) - b_0 \\ \text{s.t.} \quad & \begin{bmatrix} c_1^T \\ c_2^T \end{bmatrix} W \geq \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\ & W \geq 0 \end{aligned} \tag{8}$$

Moreover, we can compute solutions satisfying the KKT conditions and claim their optimality because g^{-1} is convex due to the concavity of g and the slater's condition is satisfied by Assumption 2 in the constraints.

Theorem 3. $p_{00} = g^{-1}(0)$*,* $p_{01} = g^{-1}(1)$*,* $p_{10} = g^{-1}(0)$*,* $p_{11} = g^{-1}(1)$ *is an optimal solution of problem* (8) *with optimal optimal value* $\gamma g(1)$*.*

Proof. Introduce λ_1, λ_2 as the Lagrangian multipliers of the two inequality constraints, and $\mu = [\mu_{00}, \mu_{01}, \mu_{10}, \mu_{11}]^T$ as the Lagrangian multipliers of the non-negativity constraints with corresponding subscripts.

We can write out the KKT conditions as follows:

$$\begin{cases} c_0 \cdot \nabla g^{-1}(W) - \lambda_1 c_1 - \lambda_2 c_2 - \mu = 0 \\ \lambda_1(c_1^T W - b_1) = 0 \\ \lambda_2(c_2^T W - b_2) = 0 \\ \mu_{ij} W_{ij} = 0 \quad \text{for } i, j = 0, 1 \\ \lambda_1, \lambda_2, \mu \geq 0 \end{cases} \tag{9}$$

Any point satisfying this KKT condition is optimal. Specifically, when $W = [0, 1, 0, 1]^T$, we have $c_1^T W - b_1 = c_2^T W - b_2 = 0$, and thus $\mu_{01} = \mu_{11} = 0$, and $\lambda_1, \lambda_2, \mu_{00}, \mu_{11} \geq 0$. Denote $c_j[k]$ as the k^{th} coordinate of vector c_j. The Lagrangian becomes:

$$\begin{bmatrix} c_1[1] & c_2[1] & 1 & 0 \\ c_1[2] & c_2[2] & 0 & 0 \\ c_1[3] & c_2[3] & 0 & 1 \\ c_1[4] & c_2[4] & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \mu_{00} \\ \mu_{01} \end{bmatrix} = \begin{bmatrix} c_0[1]\nabla g^{-1}(0) \\ c_0[2]\nabla g^{-1}(1) \\ c_0[3]\nabla g^{-1}(0) \\ c_0[4]\nabla g^{-1}(1) \end{bmatrix} \tag{10}$$

Solving (10), we have

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \mu_{00} \\ \mu_{01} \end{bmatrix} = \begin{bmatrix} \nabla g^{-1}(1)\gamma \\ 0 \\ \nabla g^{-1}(0)(1-\gamma)(1-\pi_{00}) \\ \quad + \nabla g^{-1}(1)\gamma(1-\pi_{10}) \\ \nabla g^{-1}(0)\gamma(1-\pi_{00}) \\ \quad + \nabla g^{-1}(1)\gamma\pi_{10} \end{bmatrix} \tag{11}$$

which satisfies non-negativity multiplier constraints. □

The key feature of the optimal contract for this model is that payments depend solely on the expenditure level. Low expenditures receive a payment of $g^{-1}(0)$, while high expenditures receive $g^{-1}(1)$. As a result, our approach is ineffective here because risk aversion causes providers to demand compensation for high expenditures, regardless of the outcome. Thus, it becomes impossible to incentivize providers to minimize expenditures for poor responders.

IV. NUMERICAL SIMULATION

Though intracranial pressure (ICP) monitoring is generally advised for patients with a severe traumatic brain injury (TBI), its impact on patient outcomes is not well-established. There is evidence that ICP monitoring reduces in-hospital and two-week post-injury mortality [15]. However, more recent studies question its value because of its cost-effectiveness and risk/benefit ratio [16], [17].

To numerically evaluate our optimal contracts, we first extracted a cohort of 25934 patients from the MIMIC-IV database (Medical Information Mart for Intensive Care, version 4) [18]. Out of the cohort, 728 patients were assigned ICP monitoring. The selection criteria are as follows: admitted to ICU due to traumatic brain injury or neurological disease; no relevant data is missing; has available data on the Glasgow Coma Score (GCS). The admission diagnosis is identified by the International Classification of Diseases (ICD) code in the database. For patients with multiple ICU stay records in the database, only the first ICU stay is kept.

In our framework, we aim to classify the cohort by the treatment given as well as the response level calculated in the data-driven approach illustrated in the algorithm below. The responder score is inspired by [19]. We first obtained a refined cohort using 1-1 propensity score matching and obtained 1456 patients. For the treatment and control group, we estimated a Cox proportional hazards model [20] and obtained the log hazard ratio coefficients. Assuming a common baseline hazard function, one can take the difference between treatment and control coefficients and obtain a patient-level response score to the treatment. Depending on the practical need, one can choose a cutoff point (here, we select cutoff = 0) on the response scores and cluster the patients to obtain cluster-level outcome rates.

The estimated response scores from the algorithm are shown in Fig. 1, which shows that the monitoring assignment is not strongly related to the responsiveness of the patients. This lack of reference allows for better expenditure allocation and incentivizes desirable actions. The estimated parameters are shown in Table I.

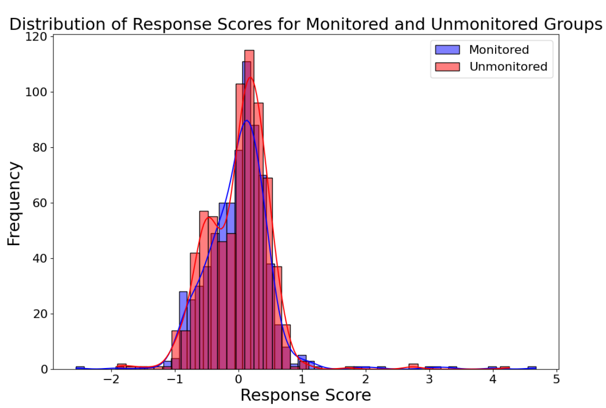

Fig. 1. Response Scores Distribution

Algorithm 1 Clustering by Treatment and Response and Calculate Outcome Rates

Require: The sample space S with the following attributes for each patient i:
1) Baseline covariates, $Z^i \in \mathbb{R}^p$
2) Propensity score, $p_i \in [0, 1]$
3) Expenditure level, $e_i \in \{0, 1\}$
4) Death in days, $t_i \in \mathbb{Z}^+$

1-1 PS matching based on p_i for every $e_i = 1$ from S and obtain refined cohort S'

For both treatment and control group in S', fit two separate Cox models

$$\log(h_0(t)) = \log(h_{00}(t)) + (\beta_{10}Z_1 + \ldots + \beta_{p0}Z_p)$$
$$\log(h_1(t)) = \log(h_{01}(t)) + (\beta_{11}Z_1 + \ldots + \beta_{p1}Z_p)$$

for each i in S' **do**
  Compute $D(Z_i) = (\beta_{11} - \beta_{10})Z_1^i + \ldots + (\beta_{p1} - \beta_{p0})Z_p^i$
end for
for each combination of k and $(r, e) \in \{0, 1\}^2$ **do**
  Extract outcome rates estimate:

$$\hat{\pi}_{r,e} = \frac{\sum_{e_i = e\ ,\ D(Z_i)^+ = r} f(t_i)}{\sum_{i \in S'} \mathbb{1}(e_i = e \text{ and } D(Z_i)^+ = r)}$$

  Where $(\cdot)^+ = \mathbb{1}(\cdot \geq 0)$ and $f(\cdot)$ is an outcome-related indicator (e.g., $f(t_i) = \mathbb{1}(t_i < los_i)$ checks if a patient dies before discharge by length of stay (los_i) in ICU).
end for
return $\hat{\pi}_{r,e}$ for each group (r, e)

TABLE I

OUTCOME PROBABILITIES BY EXPENDITURE AND RESPONDER STATUS

$\hat{\pi}$	$\hat{\pi}_{00}$	$\hat{\pi}_{01}$	$\hat{\pi}_{10}$	$\hat{\pi}_{11}$
Estimates	0.51	0.75	0.66	0.85

For estimation of F, we cite [21], where "the average cost fluctuates between €7,600 and €9,000 per hospitalization", and $F = 1$ is the normalization of this cost at around \$10,000. For the estimation of γ, we estimate it as the proportion of good responder patients with a positive response score in our dataset. $\hat{\gamma} = (335 + 309)/1456 = 0.44$.

The above estimates satisfy all model parameter assumptions for the non-negative payment model. From Section IV, if we select an optimal policy with the largest incentive gap, the optimal payment is $p_{00} = 0, p_{01} = 0, p_{10} = 0, p_{11} = \frac{1}{\hat{\pi}_{11}} = \frac{1}{0.85} \approx 1.18$, with an incentive gap $v_2 = 1 - \frac{\hat{\pi}_{01}}{\hat{\pi}_{11}} = 1 - \frac{0.75}{0.85} = 0.12$ and an objective value $\hat{\gamma} = 0.44$. Given a normalized $F = 1$, these values are percentages relative to the actual F, approximately \$10,000 in our case. Thus, the payment for high expenditure with a desirable outcome should be $\$10,000 \times 1.18 = \$11,800$, with a $\$10,000 \times 0.12 = \$1,200$ incentive gap for low expenditure with unfavorable outcomes. This results in an optimal expected payment of $\$10,000 \times 0.44 = \$4,400$.

We compare the proposed optimal contract with two extreme policies in Fig. 2: high-expenditure only and low-

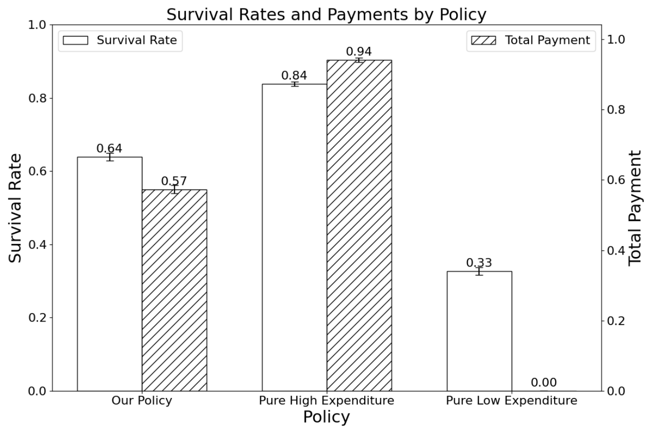

Fig. 2. Simluation Comparison with Extreme Policies

expenditure only treatment practices. In the optimal plan, where only patients with a positive response score are treated, a 64% survival rate is achieved with a 0.57 payment investment. Compared to the aggressive policy, which has an 84% survival rate with a 0.94 payment, our policy achieves a higher average outcome-rate/payment ratio ($\frac{0.64}{0.57} > \frac{0.84}{0.94}$) and a better marginal outcome-rate/payment ratio compared to the conservative policy ($\frac{0.64-0.33}{0.57} > \frac{0.84-0.33}{0.94}$), demonstrating its better cost-effectiveness.

V. CONCLUSION

Optimal contracts have been designed across three different models to ensure that payment structures incentivize providers to increase expenditure for good responders while encouraging reduced expenditure for poor responders. Numerical simulation utilizing MIMIC-IV data on ICP monitoring for TBI patients was used to evaluate a new policy. When healthcare providers are generally risk averse, our results show that in this case, the optimal contract is such that the risk aversion leads to a situation where the providers demand to be compensated when they produce high expenditures, regardless of the outcome. This means it is impossible to incentivize providers to induce low expenditures for bad-responding patients.

ACKNOWLEDGEMENTS

The authors thank Malini Mahendra, MD, in the Department of Pediatrics, Division of Pediatric Critical Care, UCSF Benioff Children's Hospital and Philip R. Lee Institute for Health Policy Studies, University of California, San Francisco, CA, USA, for her input about ICP monitoring.

REFERENCES

[1] I. Duncan, T. Ahmed, H. Dove, and T. L. Maxwell, "Medicare cost at end of life," *American Journal of Hospice and Palliative Medicine®*, vol. 36, no. 8, pp. 705–710, 2019.

[2] A. S. Iyer, C. A. Goodrich, M. T. Dransfield, S. S. Alam, C. J. Brown, C. S. Landefeld, M. A. Bakitas, and J. R. Brown, "End-of-life spending and healthcare utilization among older adults with chronic obstructive pulmonary disease," *The American journal of medicine*, vol. 133, no. 7, pp. 817–824, 2020.

[3] X. Luta, B. Ottino, P. Hall, J. Bowden, B. Wee, J. Droney, J. Riley, and J. Marti, "Evidence on the economic value of end-of-life and palliative care interventions: a narrative review of reviews," *BMC Palliative Care*, vol. 20, no. 1, pp. 1–21, 2021.

[4] R. J. Anderson, S. Bloch, M. Armstrong, P. C. Stone, and J. T. Low, "Communication between healthcare professionals and relatives of patients approaching the end-of-life: A systematic review of qualitative evidence," *Palliative medicine*, vol. 33, no. 8, pp. 926–941, 2019.

[5] L. T. Starr, C. M. Ulrich, K. L. Corey, and S. H. Meghani, "Associations among end-of-life discussions, health-care utilization, and costs in persons with advanced cancer: a systematic review," *American Journal of Hospice and Palliative Medicine®*, vol. 36, no. 10, pp. 913–926, 2019.

[6] J. C. Tangeman, C. B. Rudra, C. W. Kerr, and P. C. Grant, "A hospice-hospital partnership: Reducing hospitalization costs and 30-day readmissions among seriously ill adults," *Journal of Palliative Medicine*, vol. 17, no. 9, pp. 1005–1010, 2014.

[7] C. Donaldson and K. Gerard, "Paying general practitioners: shedding light on the review of health services." *The Journal of the Royal College of General Practitioners*, vol. 39, no. 320, pp. 114–117, 1989.

[8] R. E. Bernacki, D. N. Ko, P. Higgins, S. N. Whitlock, A. Cullinan, R. Wilson, V. Jackson, C. Dahlin, J. Abrahm, E. Mort *et al.*, "Improving access to palliative care through an innovative quality improvement initiative: an opportunity for pay-for-performance," *Journal of Palliative Medicine*, vol. 15, no. 2, pp. 192–199, 2012.

[9] J. Slutsman, L. L. Emanuel, D. Fairclough, D. Bottorff, and E. J. Emanuel, "Managing end-of-life care: Comparing the experiences of terminally ill patients in managed care and fee for service," *Journal of the American Geriatrics Society*, vol. 50, no. 12, pp. 2077–2083, 2002.

[10] Y. Lee and A. Aswani, "Optimally designing cybersecurity insurance contracts to encourage the sharing of medical data," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, 2022, pp. 6782–6787.

[11] D. Li, M. Su, X. Guo, W. Zhang, and T. Zhang, "The effect of medical choice on health costs of middle-aged and elderly patients with chronic disease: Based on principal-agent theory," *International Journal of Environmental Research and Public Health*, vol. 19, no. 13, p. 7570, 2022.

[12] B. Frandsen, M. Powell, and J. B. Rebitzer, "Sticking points: Common-agency problems and contracting in the us healthcare system," *The RAND Journal of Economics*, vol. 50, no. 2, pp. 251–285, 2019.

[13] M. Kadu, J. M. Sutherland, L. Abrahamyan, and W. P. Wodchis, "Designing financial incentives for integrated care: A case study of bundled care," *Handbook Integrated Care*, pp. 939–954, 2021.

[14] A. Sonenberg and A. L. Sepulveda-Pacsi, "Medicare payment: Advanced care planning," *The Journal for Nurse Practitioners*, vol. 14, no. 2, pp. 112–116, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1555415517309625

[15] A. Aiolfi, E. Benjamin, D. Khor, K. Inaba, L. Lam, and D. Demetriades, "Brain trauma foundation guidelines for intracranial pressure monitoring: compliance and effect on outcome," *World journal of surgery*, vol. 41, pp. 1543–1549, 2017.

[16] E. D. Mikkonen, M. B. Skrifvars, M. Reinikainen, S. Bendel, R. Laitio, S. Hoppu, T. Ala-Kokko, A. Karppinen, and R. Raj, "One-year costs of intensive care in pediatric patients with traumatic brain injury," *Journal of Neurosurgery: Pediatrics*, vol. 27, no. 1, pp. 79–86, 2020.

[17] P. Anania, D. Battaglini, J. P. Miller, A. Balestrino, A. Prior, A. D'Andrea, F. Badaloni, P. Pelosi, C. Robba, G. Zona *et al.*, "Escalation therapy in severe traumatic brain injury: how long is intracranial pressure monitoring necessary?" *Neurosurgical Review*, vol. 44, pp. 2415–2423, 2021.

[18] A. E. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. Hao, B. Moody, B. Gow *et al.*, "MIMIC-IV, a freely accessible electronic health record dataset," *Scientific data*, vol. 10, no. 1, p. 1, 2023.

[19] F. Bovis, L. Carmisciano, A. Signori, M. Pardini, J. R. Steinerman, T. Li, A. P. Tansy, and M. P. Sormani, "Defining responders to therapies by a statistical modeling approach applied to randomized clinical trial data," *BMC medicine*, vol. 17, no. 1, pp. 1–10, 2019.

[20] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, 1972.

[21] J. Berg, "Economic evidence in trauma: a review," *The European Journal of Health Economics*, vol. 5, no. Suppl 1, pp. s84–s91, 2004.