DECISION ANALYSIS
Vol. 7, No. 4, December 2010, pp. 358–365
ISSN 1545-8490 | EISSN 1545-8504 | 10 | 0704 | 0358

informs

DOI 10.1287/deca.1100.0185

Eliciting Patients' Revealed Preferences: An Inverse Markov Decision Process Approach

Zeynep Erkin
Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, zee2@pitt.edu

Matthew D. Bailey
School of Management, Bucknell University, Lewisburg, Pennsylvania 17837, matt.bailey@bucknell.edu

Lisa M. Maillart, Andrew J. Schaefer
Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261
{maillart@pitt.edu, schaefer@ie.pitt.edu}

Mark S. Roberts
Department of Health Policy and Management, University of Pittsburgh, Pittsburgh, Pennsylvania 15261,
robertsm@upmc.edu

Estimating patient preferences over various health states is an important problem in health care decision modeling. Direct approaches, which involve asking patients various abstract questions, have significant drawbacks. We propose a new approach that infers patient preferences based on observed decisions via inverse optimization techniques. We illustrate our methods on the timing of a living-donor liver transplant.

Key words: Markov decision processes; health care; inverse optimization; quality-adjusted life years
History: Received on March 22, 2010. Accepted on July 13, 2010, after 2 revisions. Published online in *Articles in Advance* September 3, 2010.

1. Motivation

Quantitative models of patient-oriented decision making require values that map a variety of health outcomes to $\mathbb{R}_+$. If the patient seeks to maximize life expectancy, then these values are simply the expected survival time associated with each health outcome. However, patients do not value every living outcome equally; "perfect health" is preferred to paralysis. The most common approach to capturing these preferences is to assign a value to each health state, known as "quality-adjusted survival" or "quality-adjusted life years" (QALYs) (Gold et al. 1997). A year in "perfect health" is worth one QALY, whereas death is given the value of zero QALYs.

It can be shown that the QALY measure does not necessarily correspond to a von Neumann–Morgenstern utility function (Torrance and Feeny 1989, Weinstein and Fineberg 1980). However, Garber and Phelps (1997) claim that QALYs well approximate utility functions. Culyer (1989) argues that the QALY measure is closer to true patient preferences than von Neumann–Morgenstern utilities and is more useful in practice. For further details on the relationship between QALYs and von Neumann–Morgenstern utility functions, see Drummond et al. (2005).

Researchers have devoted enormous effort to assess the values patients place on health states (Gold et al. 1997). The most theoretically appealing and widely applied method for these purposes is the *standard gamble* (von Neumann and Morgenstern 1947, Torrance 1976). The standard gamble ascertains the probability p at which the patient is indifferent between staying in her current health state (e.g., severe fatigue and jaundice induced by hepatitis C) for the remainder of her life, and a lottery where she moves into "perfect health" with probability p and death with probability $1-p$; the value associated with each year spent in her current state is then set equal to p years of perfect health. Another method is the *time trade-off* (Rosser and Kind 1978, Torrance et al. 1972). Under this method, the patient is asked to determine the amount of time spent in perfect health

that is equivalent to a prespecified amount of time in her current health state. The time trade-off method values the current health state as the ratio between the time spent in perfect health and that spent in the current health state.

These techniques have attracted criticism because the methods often produce inconsistent values when patients are reassessed, and, not surprisingly, different techniques often produce different values. A summary of the drawbacks and potential biases of direct preference assessment techniques is found in Gold et al. (1997), Arnold et al. (2009), and recent behavioral economics research (cf. Camerer et al. 2004 and references therein).

Faced with a similar assessment problem in the context of utility theory, Samuelson (1938, 1948) proposed that utility functions be estimated through *revealed preferences*. Rather than eliciting utility functions from consumers directly, Samuelson (1938, 1948) suggested that inferences about utility functions may be made from consumer choices. For instance, when Bundles A and B of goods are affordable, a consumer who purchases Bundle A indicates that her utility of A is at least as much as her utility of B. As such, the consumer has revealed her preference of Bundle A over B. In this manner, some of the decision maker's ordinal preferences may be observed.

We propose a similar approach to estimating patient preferences over health states based on a patient's observed behavior. We assume that a risk-neutral patient (or a physician acting on behalf of the patient) makes decisions that maximize her expected QALYs under health state valuations that are known only to herself. The assumption that a risk-neutral patient maximizes expected QALYs is common in the health state valuations assessment literature (Gold et al. 1997) and has been theoretically justified (Bleichrodt et al. 1997). The goal is to find a set of health state valuations such that the patient's observed behavior is optimal. Of course, if there exists a set of health state valuations under which the observed behavior is optimal, there are infinitely many such sets (for example, the observed behavior will be optimal under any positive scalar multiple of these valuations). To mitigate this complication, we use non-quality-adjusted expected survival as a base set of patient health state valuations. By considering restricted perturbations of these base valuations as a function of observed patient behavior, we arrive at a refined estimate of patient preferences. Such an approach can be categorized in the mathematical framework of *inverse optimization*. We caution that revealed preference approaches also appear to be subject to framing effects (Vrecko et al. 2009), which may limit the efficacy of our approach.

There are pragmatic reasons for our approach. Attempts to directly determine patient preferences are typically limited to no more than a few medical centers, and as few as several dozen patients. Using our approach, patient preferences can be assessed from deidentified data; the patient need not know that her preferences are being assessed. As such, our method can be applied to national data sets to estimate aggregate patient preferences. However, we caution that our method does not apply if limited or no data are available. In such a case, traditional methods of assessing patient preferences are the only option.

We illustrate our proposed inverse optimization technique on the optimal timing of living-donor liver transplantation. Applying the technique in this domain may support future research on how individual patients would react to changes in the national liver allocation system. Answering this question requires a model of the allocation system in which patients make accept/reject decisions when organs are offered to them, and therefore requires the specification of patients' health preferences. Recent work (Alagoz et al. 2004, 2007a, b; Sandikci et al. 2008) has analyzed this sequential decision-making problem; however, because quality-adjusted data do not exist, the valuations were not quality adjusted. Using the proposed revealed preference approach, we can more accurately parameterize the health state valuations of individual accept/reject decision models that consider patient health, organ quality, and waiting list rank.

The remainder of the paper is organized as follows. In §2, we formalize our inverse optimization approach for a generic Markov decision process (MDP). We describe a specific MDP application concerning living-donor liver transplantation and present a numerical example in §3. We conclude in §4.

2. Inverse Markov Decision Processes

An inverse optimization problem adjusts the parameters of a given optimization problem so that a particular feasible solution becomes an optimal solution. More specifically, consider an optimization problem P and a vector c. Given a feasible solution, x, and a nonnegative weight vector, w, an inverse optimization approach seeks to perturb the vector c to another vector d such that the solution x becomes an optimal solution to P with respect to the vector d and the weighted L_p norm $\|w(d-c)\|_p$ is minimized. Consistent with the existing literature, a vector for which the solution x is optimal is called *inverse feasible*.

Inverse optimization has been widely applied in various areas, such as portfolio optimization (Carr and Lovejoy 2000), transportation networks (Burton and Toint 1992, Dial 1999), and geophysical sciences (Tarantola 1987 and references therein). Ahuja and Orlin (2001) studied inverse problems under the weighted L_1 and L_∞ norms. Using duality, they established relationships between the optimization problem, P, and the inverse problem when P was a shortest-path, assignment, minimum cost flow, or minimum cut problem. A detailed survey of inverse combinatorial optimization problems is in Heuberger (2004).

In the context of an MDP, given a stationary deterministic policy π°, the inverse optimization objective is to perturb the reward vector c to a new reward vector d such that the policy π° is optimal and $\|d-c\|_p$ is minimized. Consider a discounted, infinite-horizon MDP with (finite) state space S. For every state $s \in S$, let the (finite) set of feasible actions be A_s. Furthermore, for each state–action pair, let $c(s,a)$ represent the corresponding immediate expected reward, $|c(s,a)| \le M < \infty$. A transition from state s to state j when action $a \in A_s$ is chosen occurs with probability $p(j \mid s,a)$. Given these assumptions and a discount factor $0 \le \lambda < 1$, it can be shown that there exists an optimal Markovian, deterministic, stationary policy (Bellman 1957). Note, however, that many medical decision-making problems include an absorbing state (reachable from all other states) that represents death, in which case discounting is not necessary, i.e., $\lambda = 1$ is possible. Let $v^\pi(s)$ be the total expected discounted reward under such a policy π when the system begins in state s, and similarly let v be the optimal value vector, which can be obtained by solving the optimality equations

$$v(s) = \max_{a \in A_s} \left\{ c(s,a) + \lambda \sum_{j \in S} p(j \mid s,a) v(j) \right\} \quad \text{for all } s \in S. \tag{1}$$

For finite state and action models, (1) can be recast as a linear program (D'Epenoux 1963). Let $\gamma \in \mathbb{R}_+^{|S|}$ be such that $\gamma(i) \ge 0$ and $\sum_i \gamma(i) > 0$. If we require $\sum_i \gamma(i) = 1$, we may interpret γ as an initial probability distribution over S (Puterman 1994). Then the linear program formulation of (1) is given by

$$\min_v \ \sum_{j \in S} \gamma(j) v(j) \tag{2a}$$

$$\text{subject to } \ v(s) \ge c(s,a) + \lambda \sum_{j \in S} p(j \mid s,a) v(j), \quad \forall\, s \in S,\ a \in A_s, \tag{2b}$$

$$v(s) \text{ free}, \quad \forall\, s \in S. \tag{2c}$$

Given a feasible stationary deterministic policy π° for the MDP defined by (1), we seek to perturb the reward vector c to a new set of rewards $d(s,a)$ such that the weighted L_1 norm, $\sum_{(s,a)} w(s,a) \mid d(s,a) - c(s,a)|$, is minimized and π° is optimal for the MDP defined by (1) when the vector c is replaced by the vector d. Specification of the weights, $w \in \mathbb{R}_+^{|S| \times |A|}$, is problem specific and depends on the state definition (see, for example, §3.2). The new reward vector d determined through this approach is said to be *inverse feasible* with respect to π°.

By complementary slackness, π is the optimal solution for the MDP defined by (1) if and only if it is feasible and

$$\begin{aligned} \pi(s) = a \text{ implies } & v^\pi(s) \\ & = c(s,a) - \lambda \sum_{j \in S} p(j \mid s,a) v^\pi(j), \quad \forall\, s \in S. \end{aligned} \tag{3}$$

In other words, if π is optimal for the MDP defined by (1), then the inequality constraints corresponding to the state–action pairs specified by π are satisfied as equalities. Hence we define the "*inverse MDP*" as

$$\min_d \ \sum_{s \in S} \sum_{a \in A_s} w(s,a) |d(s,a) - c(s,a)| \tag{4a}$$

$$\text{subject to } \ v(s) \ge d(s,a) + \lambda \sum_{j \in S} p(j \mid s,a) v(j), \quad \forall\, s \in S,\ a \ne \pi^\circ(s), \tag{4b}$$

$$v(s) = d(s,a) + \lambda \sum_{j \in S} p(j \mid s,a) v(j), \quad \forall s \in S,\ a = \pi^\circ(s), \tag{4c}$$

$$d \in D, \tag{4d}$$

$$v(s) \text{ free}, \quad \forall s \in S, \tag{4e}$$

$$d(s,a) \geq 0, \quad \forall s \in S,\ \forall a \in A_s, \tag{4f}$$

where the set D in (4d) is a (possibly polyhedral) set that represents additional requirements on the form of the rewards, e.g., monotonicity. By minimizing the weighted norm of the distance from a vector c, the inverse optimization problem finds the inverse-feasible vector d that is "closest" to c among all inverse-feasible vectors. Therefore, care must be given to the choice of both c and w.

Our approach makes the following assumptions:

Assumption 1. *The patient is a risk-neutral decision maker who maximizes total expected discounted reward.*

Assumption 2. *The discount rate, λ, is known for each patient.*

Assumption 3. *All decisions are based on patient physiology alone.*

Assumption 4. *The patient has complete knowledge of the transition probabilities governing disease progression.*

These assumptions ensure that the patient's decision process is well modeled by the MDP framework. If the patient's MDP model is flawed because one of these assumptions is not adequately met, then the changes made to the rewards through the inverse optimization procedure may be due to these other deficiencies rather than simply the misspecification of the rewards, and hence not reflect the true rewards.

We recognize that these assumptions may not always hold in practice. Relaxing Assumption 1 would require modeling the inverse of a risk-sensitive MDP (Howard and Matheson 1972, Jaquette 1976, Porteus 1975) and considering alternative optimality criteria. Relaxing Assumption 2 would introduce nonlinearities into the mathematical program given by (4) becoming a nonlinear program. Furthermore, adding additional parameters to be inferred (such as a patient's risk sensitivity or discount rate) would require additional terms in the objective function. Overcoming Assumptions 3 and 4 appears to be more difficult, particularly if our proposed inverse MDP approach is applied to deidentified patients.

3. Numerical Example: Living-Donor Liver Transplantation

We describe a living-donor liver transplantation application that we use to illustrate the concepts described in §§1 and 2. The problem is to determine when, as a function of health, a patient should consent to a living-donor transplant. Alagoz et al. (2004) studied this optimal stopping problem by formulating a discrete-time, infinite-horizon, discounted MDP model. Assuming that the patient is not entertaining deceased-donor organ offers, the state space S is comprised of a set of health states, $1, 2, \ldots, H$, and the absorbing death state, $H+1$. Two actions, wait (W) and transplant (T), are available at each decision epoch, and the objective is to maximize the total expected discounted life days of the patient. The decision epochs are defined as days. Let $p(s' \mid s)$ be the probability that the patient will be in health state s' at time $t+1$ given that she is in health state s at time t and the transplant does not occur. Let $c(s,T)$ be the total expected discounted posttransplant life days of the patient when the patient receives the transplant in health state s, and $c(s,W)$ be the expected immediate reward accrued in the current period when the patient chooses to wait in health state s. An optimal solution to this problem is obtained by solving the following optimality equations:

$$v(s) = \max\left\{c(s,T),\, c(s,W) + \lambda \sum_{s'=1}^{H+1} p(s' \mid s) v(s')\right\} \quad \text{for } s = 1, \ldots, H, \tag{5}$$

and $v(H+1) = 0$, where $v(s)$ is the maximum total expected discounted reward a patient in health state s can attain.

For this problem, the linear program given by (2a)–(2c) takes the following form:

$$\min_v \ \sum_{j \in S} \gamma(j) v(j) \tag{6a}$$

$$\text{subject to} \ \ v(s) \geq c(s,W) + \lambda \sum_{j \in S} p(j \mid s) v(j), \quad \forall s \in S, \tag{6b}$$

$$v(s) \geq c(s, T), \quad \forall s \in S, \tag{6c}$$

$$v(s) \text{ free}, \quad \forall s \in S. \tag{6d}$$

Thus, given a policy π°, the inverse MDP formulation of the living-donor liver transplantation problem is given by

$$\min_{d} \quad \sum_{s \in S} \sum_{a \in A_s} w(s, a) |d(s, a) - c(s, a)| \tag{7a}$$

subject to

$$v(s) = d(s, W) + \lambda \sum_{j \in S} p(j \mid s) v(j), \quad \forall s \in S,\ \pi^\circ(s) = W, \tag{7b}$$

$$v(s) \geq d(s, T), \quad \forall s \in S,\ \pi^\circ(s) = W, \tag{7c}$$

$$v(s) \geq d(s, W) + \lambda \sum_{j \in S} p(j \mid s) v(j), \quad \forall s \in S,\ \pi^\circ(s) = T, \tag{7d}$$

$$v(s) = d(s, T), \quad \forall s \in S,\ \pi^\circ(s) = T, \tag{7e}$$

$$d(s, a) \geq d(s+1, a), \quad \forall s \in S,\ a \in \{W, T\}, \tag{7f}$$

$$d(s, W) \leq 1, \quad \forall s \in S, \tag{7g}$$

$$v(s) \text{ free}, \ \forall s \in S, \tag{7h}$$

$$d(s, a) \geq 0, \quad \forall s \in S,\ \forall a \in \{W, T\}. \tag{7i}$$

The set D in constraint (4d) is given by (7f) and (7g), which ensure that as the patient's health deteriorates both the posttransplant life expectancy as well as the reward associated with waiting an additional day decrease, and that the expected reward gained while waiting one day does not exceed one day.

3.1. Estimation of Parameters and Implemented Policies

We model patient health using Model for End-stage Liver Disease (MELD) scores. MELD scores map three laboratory values to an integer between 6 (healthiest) and 40 (sickest). Because of data scarcity, we aggregate consecutive MELD scores into groups of two. The transition probabilities and posttransplant rewards are derived using two data sets, one provided by the United Network for Organ Sharing and the other by the Thomas E. Starzl Transplantation Institute at the University of Pittsburgh Medical Center. The former includes 28,717 adult end-stage liver disease patients and the latter 3,009 patients. Decision epochs correspond to days; hence, the non-quality-adjusted immediate expected rewards $c(s, W)$, which we refer to as the data-driven "wait" rewards, are set equal to 1 for all s. The health state transitions are modeled by the empiric disease-specific method of Alagoz et al. (2005). The posttransplant rewards $c(s, T)$, which we refer to as the data-driven "transplant" rewards, are calculated using the Cox proportional hazard model of Roberts et al. (2004). We refer to the optimal policy for the Alagoz et al. (2004) model under these rewards as the "suggested" policy.

We assume that the policy used by the patient was a control-limit policy with the threshold equal to the MELD score of the patient at the time of transplantation; that is, we assume that the patient's MELD score prior to transplantation was below this threshold, and transplantation was initiated the first time the MELD score met or exceeded the threshold. This assumption is mild given that MELD scores rarely jump by more than one from day to day and control-limit policies are almost always optimal in practice (Alagoz et al. 2004).

3.2. Numerical Results

Consider a 48-year old male patient with hepatitis C. According to the solution to (5), the optimal control-limit is MELD score 26. Suppose the implemented control-limit of this patient is MELD score 14; that is, the patient opts for transplantation earlier than the MDP model (5) suggests.

Table 1 includes the data-driven rewards, the suggested policy obtained by solving (5), the implemented policy, the weights used in the inverse MDP objective function, and the policy-driven rewards obtained by solving the inverse MDP. Table 2 includes the value of waiting and transplanting in each state under the data-driven and policy-driven rewards, as well as their difference.

In the example presented, we use an annual discount rate of 0.97. Furthermore, for each state–action pair we set the corresponding weight, $w(s, a)$, equal to the reciprocal of the discounted expected number of times that that state–action pair would be realized under the suggested policy, starting from the healthiest MELD score. If $w(s, a)$ is viewed as a "penalty" per unit change in $c(s, a)$ (i.e., per unit of $|d(s, a) - c(s, a)|$) that is incurred every time that state–action

Table 1 Policies, Rewards, and Weights for the Early Transplanter

MELD score	$c(s,W)$	$c(s,T)$	Suggested policy	Implemented policy	$w(s,W)$	$w(s,T)$	$d(s,W)$	$d(s,T)$
6–7	1	2,039	W	W	0.003506		0.8022	2,039
8–9	1	1,994	W	W	0.002137		0.4924	1,994
10–11	1	1,945	W	W	0.002364		0.4924	1,945
12–13	1	1,896	W	W	0.003152		0.3792	1,896
14–15	1	1,843	W	T	0.004783		0.3792	1,843
16–17	1	1,795	W	T	0.008497		0.3792	1,795
18–19	1	1,751	W	T	0.018095		0.3792	1,751
20–21	1	1,701	W	T	0.025127		0.3792	1,701
22–23	1	1,650	W	T	0.045513		0.3792	1,650
24–25	1	1,597	W	T	0.081978		0.3792	1,597
26–27	1	1,536	T	T		2.033	0.3792	1,536
28–29	1	1,491	T	T		11.36	0.3792	1,491
30–31	1	1,447	T	T		117.11	0.3792	1,447
32–33	1	1,384	T	T			0.3792	1,384
34–35	1	1,341	T	T			0.3792	1,341
36–37	1	1,283	T	T			0.3792	1,283
38–39	1	1,226	T	T			0.3792	1,226
40	1	1,172	T	T			0.3792	1,172

pair is realized, then setting the weights in this manner equates the total expected discounted penalty associated with each state–action pair and the magnitude of the change in the corresponding $c(s,a)$ value.

Consider, for example, the suggested policy reported in Table 1. The empty $w(s,a)$ entries correspond to state–action pairs that never occur under the suggested policy starting from the healthiest MELD score. Clearly, these state–action pairs include the suboptimal combinations, i.e., transplant (wait) actions for MELD scores below (at or above) 25. Additionally, because of the highly diagonal nature of the MELD score transition matrix, when starting from the healthiest MELD score it is impossible to reach MELD scores above 31 without first visiting a MELD score between 26 and 31. As a result, although it is optimal to transplant in MELD scores above 31, these state–action pairs will never occur when implementing this policy starting from the healthiest MELD score. For all of these "impossible" state–action pairs, we set $w(s,a)$ equal to an arbitrarily large value. The remaining weights are such that

$$\begin{aligned}&\frac{1}{0.003506}\cdot 1+\frac{1}{0.002137}\cdot 1+\frac{1}{0.002364}\cdot 1+\frac{1}{0.003152}\cdot 1\\&\quad+\frac{1}{0.004783}\cdot 1+\frac{1}{0.008497}\cdot 1+\frac{1}{0.018095}\cdot 1\\&\quad+\frac{1}{0.025127}\cdot 1+\frac{1}{0.045513}\cdot 1+\frac{1}{0.081978}\cdot 1\\&\quad+\frac{1}{2.033}\cdot 1{,}536+\frac{1}{11.36}\cdot 1{,}491\\&\quad+\frac{1}{117.11}\cdot 1{,}447=2{,}848.70,\end{aligned}$$

which, as expected, is the total expected discounted reward starting from the healthiest MELD score under the suggested policy as reported in the second column of Table 2.

Table 2 Values of State–Action Pairs Under the Two Sets of Rewards for the Early Transplanter

MELD score	$v_W(s)$	$v_T(s)$	$v_W(s)$	$v_T(s)$	Difference in	
	Under reward $c(s,a)$		Under reward $d(s,a)$		$v_W(s)$	$v_T(s)$
6–7	2,848.70	2,039.04	2,077.78	2,039.04	770.92	0.00
8–9	2,761.00	1,994.12	1,994.12	1,994.12	766.88	0.00
10–11	2,625.36	1,944.81	1,944.81	1,944.81	680.55	0.00
12–13	2,467.09	1,896.11	1,897.63	1,896.11	569.46	0.00
14–15	2,278.42	1,842.93	1,842.85	1,842.93	435.57	0.00
16–17	2,099.93	1,795.43	1,795.43	1,795.43	304.50	0.00
18–19	1,923.50	1,751.13	1,749.99	1,751.13	173.51	0.00
20–21	1,784.59	1,701.48	1,699.89	1,701.48	84.70	0.00
22–23	1,666.34	1,649.87	1,647.38	1,649.87	18.96	0.00
24–25	1,601.07	1,597.13	1,595.52	1,597.13	5.55	0.00
26–27	1,527.84	1,536.28	1,526.65	1,536.28	1.18	0.00
28–29	1,478.54	1,490.84	1,477.85	1,490.84	0.68	0.00
30–31	1,404.91	1,446.97	1,404.21	1,446.97	0.70	0.00
32–33	1,342.11	1,384.36	1,341.47	1,384.36	0.64	0.00
34–35	1,300.08	1,340.93	1,299.46	1,340.93	0.62	0.00
36–37	1,229.30	1,283.04	1,228.68	1,283.04	0.62	0.00
38–39	1,164.66	1,225.74	1,164.04	1,225.74	0.62	0.00
40	1,060.92	1,171.74	1,060.29	1,171.74	0.62	0.00

As seen in Table 1, the revised transplant rewards, $d(s, T)$, are identical to the data-driven rewards. However, the revised wait rewards, $d(s, W)$, exhibit a stepwise nonincreasing structure. This structure can be interpreted as a reflection of quality of life preferences across MELD scores and/or a preference to end the optimal stopping problem sooner rather than later; that is, the patient places less value on days spent in sicker states and/or places less value on days spent living with uncertainty as to when the transplant will occur. Also noteworthy is the fact that the arbitrarily large weights need not be very large to produce the same result; indeed, any value greater than approximately 0.15 for these weights produces the same vector d.

An instance for a "late transplanter," i.e., a patient who opts for transplantation later than the MDP model (5) suggested, can be structured similarly. Intuition suggests that "late transplanters" value waiting (transplanting) more (less) than is reflected by the data-driven rewards.

4. Conclusion and Future Research

Estimating patient preferences is an important problem, but traditional techniques suffer from various drawbacks, namely, the fact that it is difficult to obtain large samples, patients may find questionnaires hard to follow, and patients may provide logically inconsistent responses. We propose a new, indirect method for inferring patient preferences based on their observed policies. We formulate this problem as an inverse MDP and use linear programming to solve it. We illustrate our techniques on the problem of timing a living-donor liver transplant as a proof of concept. More realistic models that include deceased-donor liver transplantation as an alternative to the living-donor liver (Alagoz et al. 2007a, b) could also be considered with proper modifications to the inverse MDP model.

Future work could include this method's application to different clinical decisions and the use of the inferred patient preferences in societal decision models. Such a model could, for example, examine the effect of patients using the inferred patient preferences under a different liver allocation system. We also leave for future work the relaxation of the assumptions described in §2. Although relaxing Assumptions 1 and 2 appears to be possible through more difficult optimization models, Assumptions 3 and 4 may be necessary for our approach, particularly with deidentified data.

Acknowledgments

This research was supported by National Science Foundation Grants CMMI-0546960 and CMMI-0726955. The authors thank an anonymous associate editor and an anonymous referee, whose comments have improved this paper.

References

Ahuja, R. K., J. B. Orlin. 2001. Inverse optimization. *Oper. Res.* **49**(5) 771–783.

Alagoz, O., L. M. Maillart, A. J. Schaefer, M. S. Roberts. 2004. The optimal timing of living-donor liver transplantation. *Management Sci.* **50**(10) 1420–1430.

Alagoz, O., L. M. Maillart, A. J. Schaefer, M. S. Roberts. 2007a. Choosing among living donor and cadaveric livers. *Management Sci.* **53**(11) 1702–1715.

Alagoz, O., L. M. Maillart, A. J. Schaefer, M. S. Roberts. 2007b. Determining the acceptance of cadaveric livers using an implicit model of the waiting list. *Oper. Res.* **55**(1) 24–36.

Alagoz, O., C. L. Bryce, S. M. Shechter, A. J. Schaefer, C. H. Chang, D. C. Angus, M. S. Roberts. 2005. Incorporating biological natural history in simulation models: Empiric estimates of the progression of end-stage liver disease. *Medical Decision Making* **25**(6) 620–632.

Arnold, D., A. Girling, A. Stevens, R. Lilford. 2009. Comparison of direct and indirect methods of estimating health state utilities for resource allocation: Review and empirical analysis. *British Medical J.* **339** http://www.bmj.com/cgi/content/abstract/339/jul20_3/b2688.

Bellman, R. 1957. *Dynamic Programming*. Princeton University Press, Princeton, NJ.

Bleichrodt, H., P. Wakker, M. Johannesson. 1997. Characterizing QALYs by risk neutrality. *J. Risk Uncertainty* **15**(2) 107–114.

Burton, D., Ph. L. Toint. 1992. On an instance of the inverse shortest paths problem. *Math. Programming* **53**(1–3) 45–61.

Camerer, C. F., G. Loewenstein, M. Rabin. 2004. *Advances in Behavioral Economics*. Princeton University Press, Princeton, NJ.

Carr, S., W. Lovejoy. 2000. The inverse newsvendor problem: Choosing an optimal demand portfolio for capacitated resources. *Management Sci.* **46**(7) 912–927.

Culyer, A. J. 1989. The normative economics of health care finance and provision. *Oxford Rev. Econom. Policy* **5**(1) 34–58.

D'Epenoux, F. 1963. A probabilistic production and inventory problem. *Management Sci.* **10**(1) 98–108.

Dial, R. B. 1999. Minimal-revenue congestion pricing part I: A fast algorithm for the single-origin case. *Transportation Res. Part B: Methodological* **33**(3) 189–202.

Drummond, M. F., M. J. Sculpher, G. W. Torrance, B. O'Brien, G. L. Stoddart. 2005. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press, New York.

Garber, A. M., C. E. Phelps. 1997. Economic foundations of cost-effectiveness analysis. *J. Health Econom.* **16**(1) 1–31.

Gold, M. R., J. E. Siegel, L. B. Russell, M. C. Weinstein. 1996. *Cost Effectiveness in Health and Medicine*. Oxford University Press, New York.

Heuberger, C. 2004. Inverse combinatorial optimization: A survey on problems, methods, and results. *J. Combin. Optim.* **8**(3) 329–361.

Howard, R. A., J. E. Matheson. 1972. Risk sensitive Markov decision processes. *Management Sci.* **18**(7) 356–369.

Jaquette, S. C. 1976. A utility criterion for Markov decision processes. *Management Sci.* **23**(1) 43–49.

Porteus, E. L. 1975. On the optimality of structured policies in countable stage decision processes. *Management Sci.* **22**(2) 148–157.

Puterman, L. M. 1994. *Markov Decision Processes*. John Wiley and Sons, New York.

Roberts, M. S., D. C. Angus, C. L. Bryce, Z. Valenta, L. Weissfeld. 2004. Survival after liver transplantation in the United States: A disease-specific analysis of the UNOS database. *Liver Transplantation* **10**(7) 886–897.

Rosser, R., P. Kind. 1978. A scale of valuations of states of illness: Is there a social consensus? *Internat. J. Epidemiology* **7**(4) 347–358.

Samuelson, P. A. 1938. A note on the pure theory of consumer's behaviour. *Economica* **5**(17) 61–71.

Samuelson, P. A. 1948. Consumption theory in terms of revealed preference. *Economica* **15**(60) 243–253.

Sandikci, B., L. M. Maillart, A. J. Schaefer, O. Alagoz, M. S. Roberts. 2008. Estimating the patient's price of privacy in liver transplantation. *Oper. Res.* **56**(6) 1393–1410.

Tarantola, A. 1987. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier, New York.

Torrance, G. W. 1976. Social preferences for health states: An empirical evaluation of three measurement techniques. *Socio-Econom. Planning Sci.* **10**(3) 129–136.

Torrance, G. W., D. Feeny. 1989. Utilities and quality-adjusted life years. *Internat. J. Tech. Assessment Health Care* **5**(4) 559–575.

Torrance, G. W., W. H. Thomas, D. L. Sackett. 1972. A utility maximization model for evaluation of health care programs. *Health Services Res.* **7**(3) 118–133.

von Neumann, J., O. Morgenstern. 1947. *Theory of Games and Economic Behavior*, 2nd ed. Princeton University Press, Princeton, NJ.

Vrecko, D., A. Klos, T. Langer. 2009. Impact of presentation format and self-reported risk aversion on revealed skewness preferences. *Decision Anal.* **6**(2) 57–74.

Weinstein, M., H. C. Fineberg. 1980. *Clinical Decision Analysis*. W. B. Saunder, Philadelphia.