

Optimal acceptance of incompatible kidneys

Xingyu Ren, Michael C. Fu & Steven I. Marcus

To cite this article: Xingyu Ren, Michael C. Fu & Steven I. Marcus (10 Jan 2025): Optimal acceptance of incompatible kidneys, Journal of the Operational Research Society, DOI: [10.1080/01605682.2024.2444410](https://doi.org/10.1080/01605682.2024.2444410)

To link to this article: <https://doi.org/10.1080/01605682.2024.2444410>



View supplementary material [↗](#)



Published online: 10 Jan 2025.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

RESEARCH ARTICLE



Optimal acceptance of incompatible kidneys

Xingyu Ren, Michael C. Fu and Steven I. Marcus

University of Maryland, USA

ABSTRACT

Incompatibility between patient and donor is a major barrier in kidney transplantation (KT). The increasing shortage of kidney donors has driven the development of desensitization techniques to overcome this immunological challenge. Compared with compatible KT, patients undergoing incompatible KT are more likely to experience rejection, infection, malignancy, and graft loss. We study the optimal acceptance of possibly incompatible kidneys for individual end-stage kidney disease patients. To capture the effects of incompatibility, we propose a Markov Decision Process (MDP) model that explicitly includes compatibility as a state variable. The resulting higher-dimensional model makes it more challenging to analyze, but under suitable conditions, we derive structural properties including control limit-type optimal policies that are easy to compute and implement. Numerical examples illustrate the behavior of the optimal policy under different mismatch levels and highlight the importance of explicitly incorporating the incompatibility level into the acceptance decision when desensitization therapy is an option.

ARTICLE HISTORY

Received 6 March 2024
Accepted 13 December 2024

KEYWORDS

Incompatible kidney transplantation; Markov decision process; control limit policy

1. Introduction

Kidney transplantation is the organ transplantation of a donated kidney into a patient with end-stage kidney disease (ESKD). For most ESKD patients, transplantation is their best option. Compared with those undertaking dialysis treatment in their remaining lifetime, patients who undergo transplantation usually live longer and have better quality of life (Aimaretti & Arze, 2016). To receive cadaveric kidney offers, ESKD patients in the U.S. have to join the waitlist of the United Network for Organ Sharing (UNOS), which manages the Organ Procurement and Transplantation Network (OPTN). The OPTN kidney allocation system (KAS) works as follows (OPTN, 2023). Once a kidney is available, the UNOS will identify the matched candidates on the waitlist and offer the kidney to the patient with the highest priority, which is determined by the medical and listing status of both patients and donors (e.g., patient waiting time and the distance between the patient and donor). The transplant surgeon responsible for the care of the patient has a very short time to make a final decision on whether or not to accept the kidney for transplantation. If the kidney is declined, it will be offered to another eligible patient in the descending order of priority. Patients declining an offer will maintain their status on the waitlist without being penalized, and may even gain a higher priority on the waitlist in the future due to increased waiting time. Although the

number of kidney donors has been increasing, it is eclipsed by the number of newly-listed transplantation candidates. According to the OPTN database (OPTN, 2022e), there were 27,332 candidates receiving kidney transplantation during 2023, while 44,565 candidates were added to the waitlist. By the end of 2023, there were 88,667 candidates on the waitlist. Once added to the waiting list, candidates usually have a lengthy wait, with an average waiting time of 2.13 years since the new OPTN allocation policy was implemented in 2020 (OPTN, 2022a). Fewer than half of candidates eventually receive transplantation, and more than 5000 candidates die on the waitlist every year.

To improve access to kidney transplantation, modern desensitization techniques have been developed to overcome the human leukocyte antigen (HLA) and the ABO blood-type incompatibility (HLAi and ABOi), the major immunological barriers to kidney transplantation (Konvalinka & Tinckam, 2015; Rydberg et al., 2007). The HLA antigens are polymorphic proteins, and the ABO antigens consist of oligosaccharides expressed on donor kidney allograft. They are potential targets for the immune system of organ recipients. The extent of sensitization to the HLA antigens is reflected by the calculated panel reactive antibody (CPRA) value of a patient, which is the proportion of donors *expected* to have an HLA mismatch with that patient. Patients with a CPRA value greater than 0 are called sensitized. About 40% of patients on the

waitlist are sensitized (OPTN, 2022a). Compared with insensitive patients, highly sensitized patients have a much lower chance of finding a compatible donor: it may take years without a compatible donor being identified (Kuppachi & Axelrod, 2020).

To receive an incompatible kidney, a patient has to undergo desensitization therapies that aim to reduce or remove donor specific antibodies (DSA) prior to and after the transplantation. With the development of modern desensitization protocols, satisfactory outcomes have been observed in both HLAi and ABOi kidney transplantations. However, HLAi and ABOi kidney recipients show lower graft and patient survival rates compared to recipients of compatible kidneys (Kim et al., 2021; Koo & Yang, 2015; Morath et al., 2017; OPTN, 2022e). Incompatible kidney transplantation comes with distinct drawbacks, as desensitization therapies increase the risk of infection and malignancy (Clayton & Coates, 2017). Moreover, recipients of incompatible kidneys are more prone to experiencing acute or chronic rejection and early graft loss, in contrast to recipients with compatible kidneys (Ko et al., 2017; Koo & Yang, 2015). Even so, incompatible kidney transplantation still remains the best therapeutic option for patients who are highly sensitized and/or difficult to match, and in recent years, HLAi transplantation has become more common. According to the OPTN data from 2017 to 2021 (Lentine et al., 2023), more than 70% of deceased donor kidney recipients, and more than 40% of living donor kidney recipients had four or more HLA mismatches. On the other hand, ABOi kidney transplantation is less common: the current OPTN kidney allocation policy allows deceased donor ABOi transplantation only under certain circumstances, e.g., when the HLA mismatch level is zero (OPTN, 2023).

For patients with an incompatible directed living donor, they have the option of joining the Kidney Paired Donation (KPD) program to find a compatible living donor. The OPTN tracks every incompatible donor-recipient pair that registers in the KPD program, and the UNOS works with transplant centers to identify all possible matches where the donor in each pair is compatible with the recipient in another pair (Ashlagi & Roth, 2021). By a chain of exchanging donors through multiple pairs, a compatible match for all the recipients can be established. KPD is an effective means for offering better-matched organs to patients and thus reducing the kidney shortage, and the proportion of paired donations in living donor kidney transplantations has grown from 12.0% to 18.6% in the past 5 years. Furthermore, incompatible transplantation could be incorporated into the KPD program to further

augment its benefits. For example, currently, ABOi transplantation is not allowed in the KPD program (OPTN, 2023), whereas by incorporating incompatible transplantation into the KPD program, a more flexible and adaptable matching/pairing system can be established, potentially leading to a higher number of successful transplantations.

Although the kidney shortage is severe, it is reported that most cadaveric donor kidney offers are declined by at least one transplant surgeon before being accepted for transplantation (Husain et al., 2019), and 25% of offers are eventually discarded by transplant surgeons (OPTN, 2022a). Unsatisfactory organ quality accounts for most of the declined offers. Meanwhile, the number of kidney transplant recipients who experience graft failure and return on dialysis has been increasing every year (Fiorentino et al., 2021). To improve systemwide health outcomes, researchers have taken the perspective of a policy maker and modeled the organ allocation problem as a multi-class queueing problem or a sequential allocation problem (Akan et al., 2012; Ata et al., 2017; Su & Zenios, 2004, 2005, 2006; Tunç et al., 2022), focusing on the overall social welfare (e.g., the total organ usage) and the trade-off between efficiency and equity of the allocation system; as a result, simplified models are used for acceptance decisions at the patient level, for example, the multi-class queueing models of Tunç et al. (2022), Akan et al. (2012) and Su and Zenios (2005, 2006) categorize patients into different types based on their medical conditions and listing statuses, but assume their types unchanged over time and/or represent their acceptance strategy by the probability of acceptance. While this static classification or simplified strategy might be adequate for the purpose of kidney allocation, from the perspective of an individual patient, whether or not to accept a kidney offer is a critically important decision affecting their quality of life, motivating our focus on the decision-making process at the patient level.

Thus, our research models the transplant surgeon's (the decision maker) decision-making process to capture the dynamic nature of the patient state, ensuring that the decision to accept a possibly incompatible kidney offer aligns with the patient's specific circumstances. When a kidney offer arrives, the decision maker has to decide whether to accept it, depending on the current listing and medical state of the patient and the characteristics of the kidney offer, including both quality and compatibility. If the kidney is accepted, the patient will undergo transplantation; otherwise, the patient will wait for the next offered kidney and the current offer is no longer available. Our research primarily

focuses on patients on the cadaveric organ waitlist, but the same decision-making process extends to individual donor–patient pairs in the KPD program. Therefore, our research also applies to modeling and analyzing the decision-making behavior of individual pairs in the KPD program.

Current support tools for kidney transplant decision making, e.g., logistic regression models available at the website of the Scientific Registry of Transplant Recipients (SRTR, 2023), and the Transplant Models website developed by the Center for Surgical & Transplant Applied Research (C-STAR) at NYU Langone (Bae et al., 2019; C-STAR, 2013, 2018; Grams et al., 2012), only consider current characteristics of a fixed donor–recipient pair and predict outcomes of the transplant surgery, without including various uncertainties that the decision maker faces, including future patient state and future availability of kidney offers. To model the basic trade-off between waiting for a higher-quality and/or more compatible kidney and the risk of deterioration of health while waiting, we propose a Markov decision process (MDP) model to study this problem (Bertsekas, 2020).

Although similar problems have been studied in the context of liver transplantation (Alagoz et al., 2004, 2007a, 2007b, 2010; Batun et al., 2018; Kaufman et al., 2017), there are several major differences between liver and kidney transplantation. First, dialysis is an alternative option for ESKD patients, so the urgency may not be as severe (hence the patient’s remaining lifetime can be measured in months or years rather than in days), while liver transplantation is the only available therapy for end-stage liver disease (ESLD) patients. Moreover, HLA incompatibility is a major barrier in kidney transplantation, but the effect of HLA incompatibility is unclear in liver transplantation (Mahawar & Bal, 2004). ESLD patients under urgent medical conditions (i.e., those having a high model for end-stage liver disease (MELD) scores) are prioritized in the liver allocation system, while the patient sensitivity level and waiting time also play an important role in kidney allocation.

In the setting of kidney transplantation, David and Yechiali (1985); Ahn and Hornberger (1996); Bendersky and David (2016); Ren et al. (2023a) propose MDP models for the optimal acceptance of kidneys for individual ESKD patients, and a more recent article (Fan et al., 2022) studies the optimal timing to start dialysis treatment and accept a kidney offer (see Ren et al., (2023b) for a more comprehensive review of MDP models on individual patient organ acceptance decision making). Although organ quality and compatibility are key factors in kidney transplantation (Bae et al., 2019;

Koo & Yang, 2015), previous work has modeled them in an *implicit* manner and considered only one or the other of the two factors—not both simultaneously—which can lead to decisions that are sub-optimal. For example, an MDP model that takes only the compatibility into consideration is likely to reject an offer of low compatibility but high quality, which could have been the best choice for the patient, especially with the option of current desensitization therapies. Moreover, previous research has summarized all the short-term and long-term effects of the transplantation in a terminal reward, i.e., treating transplantation as a terminal state. As mentioned previously, low-quality or incompatible kidney recipients are more likely to encounter transplantation failure due to infection, rejection, and early graft loss, which would necessitate a return to dialysis and a desire for retransplantation shortly thereafter (e.g., within several months). Therefore, an MDP model explicitly modeling the retransplantation is desirable.

We propose an MDP model that incorporates the option of incompatible kidney transplantation *via* desensitization therapies. By including both the quality and the mismatch level *explicitly* as state variables in the kidney acceptance decision process, our model captures cost-benefit trade-offs between the quality and the compatibility of the kidney offer, and between waiting for a compatible kidney versus receiving an earlier transplantation but having to undergo desensitization treatment. We also explicitly model a patient who returns to being on dialysis and rejoins the waitlist for a retransplantation after experiencing an early graft loss, whereas previous work simply terminates the decision process upon organ acceptance. In particular, we model the probability of a transplantation failure to be a function of the patient state and both quality and compatibility of the donor kidney. Consequently, the state vector has a more complex correlation structure and the state dynamics are more complicated, which makes it more challenging to characterize the form of optimal policies and prove structural results.

In summary, our model takes into consideration various uncertainties and trade-offs, such that the long-term benefit of accepting or rejecting an incompatible kidney offer can be quantified, which not only provides transplant surgeons with a decision-support tool, but also promotes the utilization of incompatible kidney transplantation, which may further expand the donor pool. Moreover, our research results can be integrated into the design of both KAS and KPD policies (Ata et al., 2017; Su & Zenios, 2004, 2005, 2006; Tunç et al., 2022), where the individual patient decision-making procedure is an important component, e.g., in Tunç et al. (2022),

Su and Zenios (2004), individual patients are assumed to adopt a control limit-type policy.

To summarize, the main contributions of this article include the following:

1. In terms of modeling, our model is the first to incorporate both the quality and the compatibility of the kidney offer and to explicitly model transplantation failure and retransplantation. As a result, we are able to quantify the tradeoffs between the opportunity to get off dialysis earlier and having to undergo desensitization treatment with a higher possibility of a poor transplant outcome.
2. In terms of theory, we are able to prove desirable structural properties of the resulting more complicated MDP model under realistic conditions. In particular, we establish sufficient conditions for the existence of control limit-type optimal policies and identify situations where a control limit-type optimal policy does not exist if some condition is violated.
3. In terms of practice, preliminary numerical experiments that quantify the improved health outcomes illustrate the impact of incorporating both the quality and the compatibility of the kidney offer and allowing the option of incompatible kidney transplantation with desensitization treatment. The results indicate improvement in the order of an additional year of life expectancy for elder ESKD patients, representing a substantial gain since their expected remaining lifetime is less than five years without kidney transplantation.

The rest of the article is organized as follows: In Section 2, we formulate the individual patient kidney acceptance problem as an MDP model. In Section 3, under some intuitive assumptions, we derive structural properties including control limit-type optimal policies. In Section 4, we conduct numerical experiments to evaluate the behavior of the optimal policy under different quality and mismatch levels and illustrate the impact of incorporating compatibility and retransplantation. Section 5 concludes the article and points to future research directions. Proofs and details of parameter selection in the numerical experiments are included in Appendices A and B, respectively.

2. Model formulation

We formulate the individual patient kidney acceptance problem as a discrete-time, infinite-horizon MDP. The set of decision epochs is the natural numbers $\mathbb{N} = \{0, 1, 2, \dots\}$, where the unit could be

months (e.g., 1 month or 6 months). At each epoch $n \in \mathbb{N}$, the patient state is updated and at most one kidney offer may arrive. The decision to be made is whether to accept the offer based on the current patient state and both the quality and compatibility of the offer. If the decision maker accepts the kidney and the transplantation is a success, the decision process terminates; otherwise, the patient waits for the next offered kidney. The decision process terminates when a successful transplantation happens or the patient dies. The objective is to maximize the total reward accumulated over the entire decision process.

Remark 1. Patients typically receive their first kidney offer in about 80 days, and kidneys continue to become available approximately once a month (Husain et al., 2019). Since half of these offers don't meet the transplant center's criteria (King et al., 2022), a decision period of one to six months is reasonable, though this varies by patient characteristics. For example, unsensitized patients receive twenty times more offers than highly sensitized ones, and blood type O patients receive five times more offers than those with blood type B (see Tables A.77 and A.78 in OPTN, 2022a). Thus, decision periods should consider factors like sensitivity level, blood type, and other patient characteristics.

2.1. State of patient and kidney offer stochastic processes

The state space is $\mathcal{S} := S_H \times S_K \times S_M \cup \{P\}$. At each epoch n , the state s_n is either a triple (h_n, k_n, m_n) , or the post-transplantation state P .

- $\{h_n\}$: **Patient state.** h_n is a scalar summarizing patient listing and medical status and taking values in a finite set of positive integers $S_H = \{1, \dots, H, H+1\}$, where a larger value implies a worse patient state and $H+1$ represents death. For example, we can represent the patient's state by the estimated post-transplant survival (EPTS) score (Bae et al., 2019; OPTN, 2022c), which incorporates the patient's age, diabetes status, time on dialysis, etc., or other comprehensive indices.
- $\{k_n\}$: **Kidney offer state.** k_n represents the quality (e.g., kidney donor profile index (KDPI) score for deceased donors (OPTN, 2022d) and Live Donor KDPI (LKDPI) score for living donors (C-STAR, 2015; Massie et al., 2016), which incorporates the donor's age, height, weight, diabetes status, serum creatinine, etc.) of the donor kidney available at the current decision epoch and takes values in a finite set of positive integers $S_K = \{1, \dots, K, K+1\}$, where a larger value

implies worse quality and $K + 1$ means that *no* kidney offer is available.

- $\{m_n\}$: **Mismatch level**. m_n measures the compatibility between the patient and the kidney donor (e.g., ABO and HLA mismatch level), and takes values in a finite set $S_M = \{1, \dots, M\}$, where a larger value implies a higher mismatch level and 1 means perfect match.
- P : **Post-transplantation state**. The MDP will transition into the absorbing state P if the patient undergoes a *successful* transplantation. Without loss of generality, we may take $P \in \mathbb{Z}^3 \setminus \{S_H \times S_K \times S_M\}$, e.g., $P = (0, 0, 0)$ so that $\mathcal{S} \subset \mathbb{Z}^3$.

Remark 2. According to the current OPTN kidney allocation policy (OPTN, 2023), an important factor for allocating deceased-donor kidneys is the patient waiting time. Though the waiting time is not explicitly modeled as a state variable, we can incorporate it into the patient state. For example, the formula for computing the EPTS score includes the time on dialysis, which often coincides with the patient's waiting time (OPTN, 2022c, 2023).

Remark 3. The kidney offer state k_n only describes the donor status, while the mismatch level m_n is the compatibility of the donor-recipient pair. The patient needs to undergo desensitization treatment to accept an incompatible offer.

2.2. Action and action space

Let $\{a_n\}_{n \in \mathbb{N}}$ be the decision maker's *actions*. For each $n \in \mathbb{N}$, $a_n \in \mathcal{A} = \{W, T\}$ where

- W : reject the current offer and **wait** for one more period.
- T : accept the current offer for **transplantation**.

For any state $(h, k, m) \in S_H \times S_K \times S_M$, the feasible action set is

$$A(h, k, m) := \begin{cases} \{W, T\} & k \neq K + 1, \\ \{W\} & k = K + 1, \end{cases}$$

i.e., when a kidney offer is available, the decision maker can either reject by choosing W or accept by choosing T ; the only choice is to wait if the kidney offer is unavailable.

2.3. Dynamics

- $\mathcal{H}(j|i) := \mathbb{P}(h_{n+1} = j | h_n = i, a_n = W)$, $\mathcal{H}(\cdot|\cdot) : S_H \times S_H \mapsto [0, 1]$, is the probability that the patient state is j at epoch $n + 1$, given that the patient is in state i and the decision maker chooses W at epoch n . We set $\mathcal{H}(h|H + 1) = 0, h = 1, \dots, H$

and $\mathcal{H}(H + 1|H + 1) = 1$, i.e., “death states” $\{(H + 1, k, m), \forall k, m\}$ are absorbing.

- $\mathcal{K}(k_n|h_n), \mathcal{K}(\cdot|\cdot) : S_K \times S_H \mapsto [0, 1]$, is the probability that a kidney of state k_n is offered to a patient in state h_n . We assume that the distribution of the kidney state is only a function of the current patient state. We set $\mathcal{K}(K + 1|H + 1) = 1$, i.e., the patient doesn't receive any offer after death.
- $\mathcal{M}(m_n), \mathcal{M}(\cdot) : S_M \mapsto [0, 1]$, is the probability that the mismatch level between the patient and donor is m_n . The distribution of the mismatch level predominantly depends on the patient HLA characteristics and ABO blood type. By categorizing patients according to their HLA characteristics and blood type, we can assume that $\{m_n\}_{n \in \mathbb{N}}$ is an independent and identically distributed (i.i.d.) sequence of random variables, also independent of processes $\{h_n\}_{n \in \mathbb{N}}$ and $\{k_n\}_{n \in \mathbb{N}}$.
- $\mathcal{D}(h_n, k_n, m_n), \mathcal{D}(\cdot, \cdot, \cdot) : S_H \times S_K \times S_M \mapsto [0, 1]$, is the probability of transplantation failure for a patient in state h_n transplanted with a kidney of state k_n and mismatch level m_n . When the decision maker chooses to transplant, there are two possible outcomes: the transplantation is a success or a failure (e.g., early graft loss). With probability (w.p.) $1 - \mathcal{D}(h_n, k_n, m_n)$, the transplantation is a success, the state transitions to the absorbing state P and the decision process terminates. Otherwise, the patient returns to the waitlist for a retransplantation. For the latter case, the patient state is more likely to become worse and will evolve according to the following transition law.
- $\mathcal{Q}(j|i) := \mathbb{P}(h_{n+1} = j | h_n = i, \text{transplant fails at } n)$, $\mathcal{Q}(\cdot|\cdot) : S_H \times S_H \mapsto [0, 1]$, is the probability that the patient's state is j at time $n + 1$, given that the patient is in state i and a transplantation failure happens at time n . We take the function \mathcal{Q} as different from the function \mathcal{H} , considering that desensitization treatment and transplantation failure may have a negative impact on the patient's health.

Remark 4. Patient state h_n is a comprehensive index computed by the medical and listing status of the patient, which also determines the priority ranking of the patient on the waitlist. For example, as mentioned in Remark 2, we may define the patient state by the EPTS score, which is strongly correlated with the patient waiting time, an essential factor of the priority ranking. Assuming the distribution of the kidney state to be a function of the patient state implies that the chance to get a kidney offer depends on the priority ranking on the waitlist, which implicitly models the effect of the kidney allocation system.

Remark 5. To justify the assumption that $\{m_n\}_{n \in \mathbb{N}}$ is independent of $\{h_n\}_{n \in \mathbb{N}}$ and $\{k_n\}_{n \in \mathbb{N}}$, for illustration, we take the HLA mismatch level (as the current U.S. kidney allocation system matches only ABO blood-type compatible pairs), which is based solely on genetic differences between the donor and recipient. Since a patient's HLA antigen type is fixed and unrelated to their health, it is reasonable to assume that the mismatch level is independent of the patient state transition.

When a donor organ becomes available in the allocation system, for a specific patient, the distribution of mismatch level between the patient and donor is determined solely by the population's HLA antigen distribution, independent of patient state or organ quality. The allocation score, which considers factors like organ quality, patient state, and mismatch level, prioritizes patients with the highest scores. However, the score only considers HLA mismatch levels of zero or one, ignoring mismatch levels of two or higher (OPTN, 2023). With 95% of U.S. transplantations having a mismatch level of two or above (OPTN, 2022a), the mismatch level has a negligible effect on both the likelihood of receiving a kidney offer and the quality of the kidney offer in most cases. Thus, assuming that “compatibility is independent of the patient state and organ quality” is a reasonable approximation.

To summarize, we provide the overall decision-making procedure in Algorithm 1, and the general transition probability as follows: for any $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}(s_{n+1} = P | s_n = s, a_n = a) &= \begin{cases} 1 & \text{if } s = P, \\ 1 - \mathcal{D}(h, k, m) & \text{if } s = (h, k, m), a = T, \\ 0 & \text{otherwise;} \end{cases} \\ \mathbb{P}(s_{n+1} = (h', k', m') | s_n = s, a_n = a) &= \begin{cases} \mathcal{M}(m') \mathcal{K}(k' | h') \mathcal{H}(h' | h) & \text{if } s = (h, k, m), a = W, \\ \mathcal{D}(h, k, m) \mathcal{M}(m') \mathcal{K}(k' | h') \mathcal{Q}(h' | h) & \text{if } s = (h, k, m), a = T, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Algorithm 1. Kidney transplantation decision-making procedure.

Initial patient health h_1 ;
for decision epoch $n = 1, 2, \dots$ **do**
 Receive a kidney offer (k_n, m_n) w.p.
 $\mathcal{K}(k_n | h_n) \mathcal{M}(m_n)$;
 Make a decision whether to accept the offer;
 if the offer is accepted **then**
 if transplantation is successful **then**
 Receive terminal reward $r(h_n, k_n, m_n)$;
 The process **terminates**;
 else
 The transplantation fails;
 Receive reward $c(h_n)$;

The patient's state transitions to h_{n+1}
 w.p. $\mathcal{Q}(h_{n+1} | h_n)$;
if the offer is declined **then**
 Receive reward $c(h_n)$;
 The patient state transitions to h_{n+1}
 w.p. $\mathcal{H}(h_{n+1} | h_n)$;
if the patient dies, i.e. $h_{n+1} = H + 1$ **then**
 The process **terminates**;

2.4. Reward functions

- $c(\cdot) : S_H \mapsto \mathbb{R}_+$, **the intermediate reward function.** If a patient in state h_n does not undergo a successful transplantation, i.e., the decision maker either chooses W or chooses T but the transplantation fails, they get an intermediate reward $c(h_n)$ for being alive for one period. We set $c(H + 1) = 0$.
- $r(\cdot, \cdot, \cdot) : S_H \times S_K \times S_M \mapsto \mathbb{R}_+$, **the terminal reward function.** If a patient in state h_n undergoes a successful transplantation with a kidney of quality k_n and mismatch level m_n , they receive a terminal reward $r(h_n, k_n, m_n)$. Reward function r measures the long-term effect of a successful transplantation which terminates the decision process. We set $r(H + 1, k, m) = 0, \forall k \in S_K, m \in S_M$.

We assume that the reward accruing in the post-transplantation state P is zero. For any feasible state-action pair (s, a) , if $a = T$, the one-stage reward $g(s, a)$ is given by

$$g((h, k, m), T) = \begin{cases} r(h, k, m) & \text{w.p. } (1 - \mathcal{D}(h, k, m)), \\ c(h) & \text{w.p. } \mathcal{D}(h, k, m). \end{cases}$$

If $a = W$, the one-stage reward $g(s, W) = c(h)$ is a constant.

Remark 6. The death of the patient also terminates the decision process, since death states are absorbing and the patient receives zero reward upon death.

2.5. Objective function

The goal is to find a policy $\pi : S \mapsto \mathcal{A}$ that maximizes the expected total discounted reward

$$f_\pi(h, k, m) := \mathbb{E} \left(\sum_{i=0}^{\infty} \lambda^i g(s_i, \pi(s_i)) | s_0 = (h, k, m) \right)$$

for any initial state $s_0 = (h, k, k) \in S_H \times S_K \times S_M$, where $\lambda \in [0, 1]$ is the discount factor. We only consider stationary policies (i.e., policies that don't depend explicitly on time) in this article. Denote the maximum expected total discounted reward (also known as the value function) by $V(h, k, m) := \max_{\pi \in \Pi} f_\pi(h, k, m), \forall h, k$, and m , where Π is the set of stationary policies.

3. Structural results

In this section, we present two types of structural results: monotonicity of the value function $V(h, k, m)$, and the existence of control limit-type optimal policies, which will be formally defined. First, we present [Theorem 1](#) on the Bellman equation and convergence of the value iteration algorithm, which are standard MDP results and useful for computing the optimal policy and proving other structural results. As the MDP model has finite state and action spaces and is time-homogeneous, [Theorem 1](#) follows from Proposition 5.4.1 in Bertsekas (2020).

Theorem 1. *The value function V satisfies: $\forall h \in S_H, m \in S_M$, and $k < K + 1$,*

$$\begin{aligned}
 & V(h, k, m) \\
 &= \max \left(\begin{array}{l} (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)(c(h) + \lambda \sum_{h' \in S_H} v(h') \mathcal{Q}(h'|h)), \\ c(h) + \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h) \end{array} \right), \quad (1) \\
 & V(h, K + 1, m) = c(h) + \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h),
 \end{aligned}$$

where

$$v(h) = \sum_{k \in S_K} \left(\sum_{m \in S_M} \mathcal{M}(m) V(h, k, m) \right) \mathcal{K}(k|h), h \in S_H.$$

Note that $V(H + 1, k, m) = 0, \forall k \in S_K, m \in S_M$. Moreover, the sequence of functions $\{V_n\}_{n \geq 0}$ recursively defined by the value iteration procedure given by

$$\begin{aligned}
 & V_{n+1}(h, k, m) \\
 &= \max \left(\begin{array}{l} (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)(c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h)), \\ c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h) \end{array} \right), \quad (2) \\
 & V_{n+1}(h, K + 1, m) = c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h),
 \end{aligned}$$

$\forall h \in S_H, m \in S_M, k < K + 1$, converges pointwise to V , starting from any bounded function V_0 , where

$$v_n(h) = \sum_{k \in S_K} \left(\sum_{m \in S_M} \mathcal{M}(m) V_n(h, k, m) \right) \mathcal{K}(k|h), h \in S_H.$$

$v(h)$ can be interpreted as the expected total discounted reward when the patient state is h . Note that the optimal policy may not be unique, and denote $A^*(h, k, m)$ the set of optimal actions at the state (h, k, m) .

[Theorem 2](#) provides sufficient conditions to guarantee that the value function $V(h, k, m)$ is nonincreasing in both h and k . [Theorem 2](#) is intuitive: the patient overall benefit won't increase if the quality of the kidney offered or the patient's state gets worse. First, we provide several intuitive assumptions and a preliminary result that are needed to establish [Theorem 2](#).

Assumption 1. $r(h, k, m)$ is nonincreasing in any component with the other two fixed.

Throughout the article, we say that a multivariable function is monotone in some component if the function is monotone in that component with other components fixed. For example, $r(h, k, m)$ is nonincreasing in h , k , and m . [Assumption 1](#) has an

intuitive explanation that the reward for a successful transplantation does not increase if the patient state deteriorates and/or the kidney quality gets worse and/or the mismatch level increases.

Assumption 2. $c(h)$ is nonincreasing in h .

[Assumption 2](#) is also intuitive: the intermediate reward for waiting does not increase if the patient state deteriorates.

Assumption 3. $\mathcal{D}(h, k, m)$ is nondecreasing in any component with the other two fixed.

[Assumption 3](#) can be interpreted as meaning that the probability of transplantation failure does not decrease if the patient state deteriorates and/or the kidney quality gets worse and/or the mismatch level increases.

Definition 1. For a time-homogeneous discrete-time Markov chain with state space $S = \{1, \dots, n\}$, its

transition probability function $P(\cdot|\cdot) : S \times S \mapsto [0, 1]$ is stochastically increasing (Smith & McCardle, 2002) if the sequence of random variables with distribution functions $\{P(\cdot|k)\}_{k=1, \dots, n}$ is in increasing stochastic order, which is defined as follows (Ross, 1996): for random variables X and Y , we say that $X \geq_{st} Y$ if $\mathbb{P}(X > t) \geq \mathbb{P}(Y > t), \forall t$.

Specifically, $P(\cdot|\cdot)$ is stochastically increasing if $\sum_{j=k}^n P(j|i)$ is nondecreasing in i for any $k = 1, \dots, n$, where $P(j|i)$ is the transition probability from state i to j .

Assumption 4. Transition probability functions \mathcal{H} and \mathcal{Q} are stochastically increasing.

Some papers, e.g., Alagoz et al. (2004, 2007a, 2007b), refer to the stochastically increasing property as increasing failure rate (IFR). The stochastically increasing property has an intuitive explanation in the context of disease progression: the worse the patient state, the more likely the patient state is to become even worse.

Definition 2. For a time-homogeneous discrete-time Markov chain with state space $S = \{1, \dots, n\}$, we say that transition probability function $P(\cdot|\cdot) : S \times S \mapsto [0, 1]$ is stochastically greater than $Q(\cdot|\cdot) : S \times S \mapsto [0, 1]$, denoted by $P \geq_{st} Q$, if

$$\sum_{j=k}^n P(j|i) \geq \sum_{j=k}^n Q(j|i), \forall i, k \in S.$$

Assumption 5. $\mathcal{Q} \geq_{st} \mathcal{H}$.

Note that larger patient state represents worse status. Assumption 5 implies that the patient state is less likely to become worse under transition function \mathcal{H} , compared with transition function \mathcal{Q} . Assumption 5 captures the negative impact of the transplantation failure: the patient state is more likely to become worse if they experience a failed transplantation.

Assumption 6. For $k = 1, \dots, K, h = 1, \dots, H$,

$$\mathcal{K}(k|h+1) \leq \mathcal{K}(k|h). \quad (3)$$

Assumption 6 posits that patients in better states are more likely to receive an offer, aligned with the current kidney allocation rule. For instance, patients with top 20% EPTS scores are favored in the allocation of high-quality kidneys (i.e., those with KDPI less than 35.), and prioritization is given to patients who have longer waiting time. For more details, refer to Tables 7 and 8 and Sections 8.2–8.4 in OPTN (2023). This differs from liver allocation, where patients in critical medical condition (i.e., those having high model for end-stage liver disease (MELD) scores) are prioritized.

Theorem 2. Under Assumptions 1 through 6, the following hold:

1. $v(h)$ is nonincreasing in h ;
2. $V(h, k, m)$ is nonincreasing in k ;
3. $V(h, k, m)$ is nonincreasing in h .

Remark 7. Smith and McCardle (2002) proves that the value function is decreasing when the one-stage reward function is decreasing for each action and the overall transition probability of the MDP is stochastically increasing (suitably defined for the partially ordered state space), for which Assumptions 3 through 6 are neither necessary nor sufficient.

Next, we will establish sufficient conditions to guarantee the existence of control limit-type optimal policies. First, we formally define a control limit policy.

Definition 3. Consider an MDP model with one-dimensional state space $S \subset \mathbb{R}$ and an action space \mathcal{A} . A policy $\pi : S \mapsto \mathcal{A}$ is called a control limit policy if there exists a finite collection of intervals $\{I_i\}_{i=1}^n$ partitioning \mathbb{R} and satisfying

1. For any $a \in \mathcal{A}$, there exists at most one interval I_i satisfying $\pi(s) = a, \forall s \in I_i \cap S$;
2. For any interval I_i , there exists $a \in \mathcal{A}$ such that $\pi(s) = a, \forall s \in I_i \cap S$.

Endpoints of intervals $\{I_i\}_{i=1}^n$ are called control limits.

The simplest form of control limit policy is the following, which partitions the state space into two regions:

$$\pi(s) = \begin{cases} a_1 & \text{if } s < s^*, \\ a_2 & \text{if } s \geq s^*. \end{cases} \quad (4)$$

The action to take depends only on whether the state s is greater than or less than the control limit s^* , and solving the MDP problem boils down to finding the optimal threshold. Our setting necessitates a three-dimensional state vector, which requires an adjustment to Definition 3. Under suitable conditions, by fixing values of two state variables and projecting the state space onto the other dimension, we can establish optimal policies that take the form of a control limit policy (in one-dimension).

Definition 4. A policy $\pi : S_H \times S_K \times S_M \mapsto \mathcal{A}$ is called a patient-based control limit policy if there exists a control limit function $H(k, m)$ such that for each $k \in S_K, m \in S_M$, $\pi(h, k, m) = T$ if and only if the patient state $h > H(k, m)$ (or $h < H(k, m)$).

Definition 5. A policy $\pi : S_H \times S_K \times S_M \mapsto \mathcal{A}$ is called a kidney-based control limit policy if there

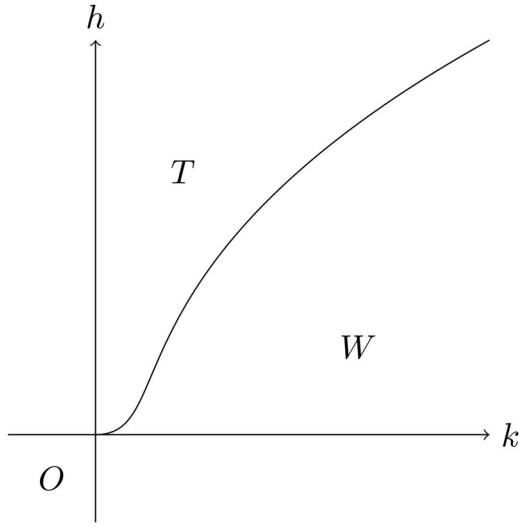


Figure 1. For an MDP with state space $S_H \times S_K$, an optimal policy such that states for which optimal actions are W and T are contained in two disjoint connected subsets.

exists a control limit function $K(h, m)$ such that for each $h \in S_H, m \in S_M$, $\pi(h, k, m) = T$ if and only if the kidney state $k < K(h, m)$ (or $k > K(h, m)$).

Definition 6. A policy $\pi : S_H \times S_K \times S_M \mapsto \mathcal{A}$ is called a match-based control limit policy if there exists a control limit function $M(h, k)$ such that for each $h \in S_H, k \in S_K$, $\pi(h, k, m) = T$ if and only if the mismatch level $m < M(h, k)$ (or $m > M(h, k)$).

Remark 8. For an MDP model with a two-dimensional state space $S_H \times S_K$, if both patient-based and kidney-based control limit optimal policies exist, it is easy to show that there exists an optimal policy such that states for which optimal actions are W and T , respectively, are contained in two disjoint connected subsets of \mathbb{R}_+^2 , as illustrated in Figure 1. There is only one decision boundary to determine. However, our MDP model has a three-dimensional state space. Existence of all three types of control limit policies does not guarantee that there exists an optimal policy that allows \mathbb{R}_+^3 to be partitioned into two connected decision regions. There are additional decision boundaries to identify, making both the computation and implementation of the policy more complicated. A counterexample is provided in Figure 2. Ren et al. (2023b) provide an example showing that the expansion of the action space has a similar effect. In general, the optimal policies may take a more complicated form if the size of either the state space or the action space or both increase.

We shall see later that all three types of control limit optimal policies are equivalent if all of them exist, i.e., once we derive one control limit function, the other two can be obtained by taking the inverse of the derived one.

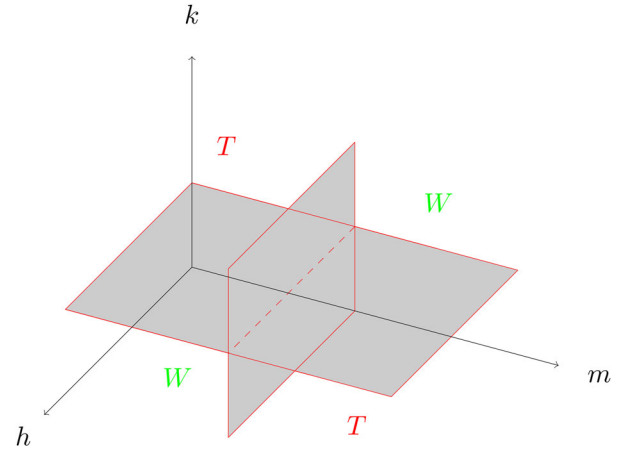


Figure 2. For an MDP with state space $S_H \times S_K \times S_M$, suppose that its unique optimal policy partitions \mathbb{R}^3 into four disjoint decision regions by a vertical plane and a horizontal plane. Both actions W and T are optimal over two disconnected regions. However, all three types of control limit optimal policies exist.

The control limit-type policy is a specific case of a monotone policy. In their work, Serfozo (1976) establishes sufficient conditions for the existence of a monotone optimal policy within the context of a discrete-time MDP, where the state space S is assumed to be partially ordered, and the action space A is assumed to be a compact subset of the real line. Define the Q-function

$$\begin{aligned} \bar{Q}(h, k, m, a) &= \begin{cases} (1 - \mathcal{D}(h, k, m))r(h, k, m) \\ + \mathcal{D}(h, k, m)(c(h) + \lambda \sum_{h' \in S_H} v(h') \mathcal{Q}(h'|h)) & \text{if } a = T, \\ c(h) + \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h) & \text{if } a = W, \end{cases} \end{aligned} \quad (5)$$

Serfozo (1976) proves the existence of a monotone optimal policy when the Q-function is submodular on $S \times A$, where submodularity is defined as follows.

Definition 7. Let X and Y be partially ordered sets and $f(x, y)$ be a real-valued function on $X \times Y$. We say that f is submodular if for $x_1 \geq x_2$ in X and $y_1 \geq y_2$ in Y ,

$$f(x_1, y_1) + f(x_2, y_2) \leq f(x_1, y_2) + f(x_2, y_1).$$

Applying the result of Serfozo (1976), it is evident that by selecting any state variable and fixing the other two, if the Q-function is submodular as a function of that variable and the action, a control limit-type optimal policy for that variable can be established (although the action space is unordered, we could artificially assign an order to it, such as $T > W$ or $T < W$). Specifically,

- The patient-based control limit optimal policy exists if $\bar{Q}(h+1, k, m, T) - \bar{Q}(h, k, m, T) > \bar{Q}(h+1, k, m, W) - \bar{Q}(h, k, m, W), \forall h$.

- The kidney-based control limit optimal policy exists if $\bar{Q}(h, k+1, m, T) - \bar{Q}(h, k, m, T) < \bar{Q}(h, k+1, m, W) - \bar{Q}(h, k, m, W), \forall k$.
- The match-based control limit optimal policy exists if $\bar{Q}(h, k, m+1, T) - \bar{Q}(h, k, m, T) < \bar{Q}(h, k, m+1, W) - \bar{Q}(h, k, m, W), \forall m$.

The first condition implies that for a patient-based control limit policy to be optimal, as a patient's health deteriorates (i.e., as h increases), the reduction in benefit from waiting must exceed the reduction in benefit from immediate transplantation. In practice, this situation may arise when a patient is in poor health, where the decline in quality of life on dialysis or the increased risk of death during waiting makes immediate transplantation more preferable as the patient state worsens. The other two conditions can be explained in a similar manner. [Theorems 3](#) through [5](#) will provide sufficient conditions to establish the aforementioned conditions, respectively.

With [Assumptions 1](#) through [6](#), we are able to prove [Theorems 3](#) and [4](#), which establish match-based and kidney-based control limit optimal policies, respectively.

Theorem 3. *Under [Assumptions 1](#) through [6](#), there exists a match-based control limit optimal policy, i.e., there exists a control limit function $M^*(h, k)$, such that for any fixed h and k , it is optimal to accept the kidney offer if and only if the mismatch level $m < M^*(h, k)$*

Theorem 4. *Under [Assumptions 1](#) through [6](#), there exists a kidney-based control limit optimal policy, i.e., there exists a control limit function $K^*(h, m)$, such that for any fixed h and m , it is optimal to accept the kidney offer if and only if the kidney state $k < K^*(h, m)$.*

Remark 9. Although the optimal policy may not be unique, “if and only if” in the theorem guarantees that there exists a unique match-based (and kidney-based) control limit optimal policy.

Using [Theorem 3](#), it can be easily shown that $V(h, k, m)$ is nonincreasing in m , i.e., the value function doesn't increase if the mismatch level increases.

Corollary 1. *Under [Assumptions 1](#) through [6](#), $V(h, k, m)$ is nonincreasing in m .*

Both [Theorems 3](#) and [4](#) are intuitive: the decision maker should accept kidney offers that are of sufficiently good quality and/or low mismatch level. Once we establish both control limit optimal policies, we can easily show the (partial) monotonicity, as well as invertibility of control limit functions.

Corollary 2. *Under [Assumptions 1](#) through [6](#), the following hold:*

1. $K^*(h, m)$ is nonincreasing in m .
2. $M^*(h, k)$ is nonincreasing in k .
3. Define $K^{-M}(h, k) := \min\{m \in S_M \mid k \geq K^*(h, m)\}$. Then, $M^*(h, k) = K^{-M}(h, k)$.
4. Define $M^{-K}(h, m) := \min\{k \in S_K \mid m \geq M^*(h, k)\}$. Then, $K^*(h, m) = M^{-K}(h, m)$.

To show [Theorem 5](#), the existence of a patient-based control limit optimal policy, we need several additional assumptions.

Assumption 7. For $h = 1, \dots, H$ and $h_0 = h + 1, \dots, H$,

$$\sum_{h'=h_0}^H \mathcal{H}(h'|h) \leq \sum_{h'=h_0}^H \mathcal{H}(h'|h+1). \quad (6)$$

The interpretation of [Assumption 7](#) is similar to the stochastically increasing property, but [Assumption 7](#) is neither a sufficient nor a necessary condition for the stochastically increasing property of \mathcal{H} .

Assumption 8. For $h = 1, \dots, H-1$, $k \leq K$, $m \in S_M$

$$\begin{aligned} & \frac{\mathbb{E}g((h, k, m), T) - \mathbb{E}g((h+1, k, m), T)}{\mathbb{E}g((h+1, k, m), T)} \\ & \leq (1 - \mathcal{D}(h, k, m))\lambda(\mathcal{H}(H+1|h+1) - \mathcal{H}(H+1|h)). \end{aligned} \quad (7)$$

[Assumption 8](#) has an intuitive explanation that, as the patient state becomes worse, the increment of the probability of death during waiting is greater than the marginal reduction in the expectation of the one-step reward for choosing T . Conditions similar to [Assumption 8](#) have been verified using real data in liver transplant studies (Alagoz et al., 2004, 2007b).

Assumption 9. $\mathcal{Q} - \mathcal{H}$ is stochastically decreasing, i.e., for $h_0 = 1, \dots, H+1$ and $h = 1, \dots, H$,

$$\begin{aligned} & \sum_{h'=h_0}^{H+1} \mathcal{Q}(h'|h+1) - \mathcal{H}(h'|h+1) \\ & \leq \sum_{h'=h_0}^{H+1} \mathcal{Q}(h'|h) - \mathcal{H}(h'|h). \end{aligned} \quad (8)$$

[Assumption 9](#) states that as patient state h increases, the “distance” between distributions $\mathcal{H}(\cdot|h)$ and $\mathcal{Q}(\cdot|h)$ decreases. In particular, [Assumption 9](#) implies that transplantation failure has a larger impact on a healthier patient, e.g., an unsuccessful transplant will result in a substantial increase in EPTS score for a low-EPTS patient, as shown in [Tables B1](#) and [B2](#).

Theorem 5. Under Assumptions 1 through 9, there exists a patient-based control limit optimal policy, i.e., there exists an optimal control limit function $H^*(k, m)$, such that for any fixed k and m , it is optimal to accept the kidney offer if and only if the patient state $h > H^*(k, m)$.

Theorem 5 has an intuitive explanation: the decision maker should accept a kidney offer if the patient state is worse than some threshold. With Theorem 5, we are able to derive more monotonicity and invertibility results of control limit functions similar to Corollary 2.

Remark 10. Under appropriate conditions, the existence of a monotone optimal policy has been established for various MDP types, including regular MDPs (Flores-Hernández & Montes-de Oca, 2007; Puterman, 2014; Serfozo, 1976), risk-sensitive MDPs (Avila-Godoy & Fernández-Gauchera, 1998), and partially observed MDPs (Lovejoy, 1987; Miehlung & Teneketzis, 2020). These conditions typically involve assumptions about the stochastic monotonicity of the overall transition probability and the submodularity of the one-period reward function. In contrast, Theorems 3 through 5 require only stochastic monotonicity in each of the individual dimensions of the state transition, aligning with the metatheorems in Oh and Özer (2016), which provide a framework for establishing threshold policies in optimal stopping problems with partially monotone state transitions.

Corollary 3. Under Assumptions 1 through 9, the following hold:

1. $K^*(h, m)$ is nondecreasing in h .
2. $M^*(h, k)$ is nondecreasing in h .
3. $H^*(k, m)$ is nondecreasing in k and m .
4. Define $K^{-H}(k, m) := \max\{h \in S_H \mid k \geq K^*(h, m)\}$. Then, $H^*(k, m) = K^{-H}(k, m)$.
5. Define $M^{-H}(k, m) := \max\{h \in S_H \mid m \geq M^*(h, k)\}$. Then, $H^*(k, m) = M^{-H}(k, m)$.
6. Define $H^{-M}(h, k) := \min\{m \in S_M \mid h \leq H^*(k, m)\}$. Then, $M^*(h, k) = H^{-M}(h, k)$.
7. Define $H^{-K}(h, m) := \min\{k \in S_K \mid h \leq H^*(k, m)\}$. Then, $K^*(h, m) = H^{-K}(h, m)$.

The proof of Corollary 3 is omitted, as it is straightforward and exactly the same as that of Corollary 2. Corollary 3 together with Corollary 2 shows that the three types of control limit optimal policies are equivalent to each other: if all three types of control limit optimal policies exist, once we obtain any one of them, we can easily obtain the other two by invertibility.

Theorem 6 considers two patients with identical state transition probability functions but different probabilities for kidney offers. For example, compared with insensitive patients, it is much harder for highly

sensitized patients to find a compatible donor. If patient 1 has a higher chance to receive a kidney offer than patient 2, then the value function of patient 1 dominates the value function of patient 2. Recall that a larger value of k indicates worse quality and $k = K + 1$ means no kidney offer is available.

Theorem 6. Let Π_1 and Π_2 be two MDPs, where the distributions of the kidney state are \mathcal{K}_1 and \mathcal{K}_2 , respectively. Suppose that $\mathcal{K}_2 \geq_{st} \mathcal{K}_1$. Let V^1 and V^2 be the value functions of Π_1 and Π_2 , respectively. If Π_1 and Π_2 have the same reward functions c and r , probability function of transplantation failure \mathcal{D} , pmf of mismatch level \mathcal{M} , and patient state transition probability functions \mathcal{H} and \mathcal{Q} , then $V^1(h, k, m) \geq V^2(h, k, m)$ for all $h \in S_H, k \in S_K, m \in S_M$.

Theorem 7 provides a similar result, which considers two patients with identical probabilities for kidney offers but different patient state transition probability functions. If the health of patient 1 is less likely to become worse, then the value function of patient 1 dominates the value function of patient 2. Recall that a larger value of h implies worse patient state and $h = H + 1$ represents death.

Theorem 7. Let Π_1 and Π_2 be two MDPs with patient state transition probability functions $(\mathcal{H}_1, \mathcal{Q}_1)$ and $(\mathcal{H}_2, \mathcal{Q}_2)$, respectively. Suppose that $\mathcal{H}_2 \geq_{st} \mathcal{H}_1$, $\mathcal{Q}_2 \geq_{st} \mathcal{Q}_1$. Let V^1 and V^2 be value functions of Π_1 and Π_2 , respectively. If Π_1 and Π_2 have the same reward functions c and r , probability function of transplantation failure \mathcal{D} , pmf of mismatch level \mathcal{M} , and distribution of kidney state \mathcal{K} , then $V^1(h, k, m) \geq V^2(h, k, m)$ for all $h \in S_H, k \in S_K, m \in S_M$.

4. Numerical experiments

In this section, we use numerical experiments to demonstrate the importance of modeling both kidney quality and compatibility and the impact of allowing the desensitization therapy option. We set parameters based on a recent OPTN data report (OPTN, 2022a). Next, we compute the optimal policy derived in Section 3 and show how its behavior and performance vary under different kidney quality and mismatch levels. Then, we compare the optimal policy with another policy that doesn't include compatibility as a state variable or retransplantation. In particular, the optimal policy performs much better at low mismatch levels. Moreover, we provide an example where Assumption 8, part of the sufficient condition for Theorem 5 to hold, is violated and the patient-based control limit optimal policy doesn't exist. We observe that retransplantation is rare under these parameter settings due to the low likelihood of receiving a kidney offer and transplantation failure. To evaluate the effects of compatibility and retransplantation

separately, we increase the chances of receiving a kidney offer and transplantation failure, and conduct an additional experiment where the mismatch level is hidden, but retransplantation is included.

We consider a 70-year old patient who starts dialysis at the beginning of the decision process and doesn't have diabetes or a prior organ transplant. We begin by describing how state variables, state space and reward functions are defined. The *decision period* is six months. The *patient state* h , taking values in $S_H = \{1, \dots, 17\}$, is defined based on the EPTS score (C-STAR, 2018; OPTN, 2022c), one of the most commonly used measures in evaluating the patient expected post-transplantation survival time. The patient state transition law is straightforward, because the EPTS score is a function of age once the patient's diabetes state, time to start dialysis, and number of prior organ transplantations are specified. We only consider deceased kidney donors and use the KDPI score to represent the *kidney state* k . Both EPTS and KDPI scores are used in the current kidney allocation system. Since a recent OPTN data report (OPTN, 2022a) categorizes kidneys into four groups based on their KDPI ranges, we define four types of kidneys, i.e., $S_k = \{1, \dots, 5\}$ where $k = 5$ represents that the kidney offer is not available. We use the degree of HLA mismatch to represent the *mismatch level* m . The number of HLA antigen mismatches of a donor-recipient pair ranges from 0 to 6, so we set $S_M = \{1, \dots, 7\}$. We assume that kidney states $\{k_n\}_{n \geq 1}$ form an i.i.d. sequence of random variables, independent of both patient state and mismatch level. If the decision maker chooses to wait, the patient accrues an intermediate reward for being alive for half a year (the decision period), i.e., $c(h) = 0.5, \forall h$. If the patient undergoes a successful transplantation, they receive terminal reward $r(h, k, m)$ equal to the expected post-transplant lifetime.

We consider three experiments. The first setting models many of the parameter values for a typical 70-year old patient with ESKD. The second setting is similar, with the exception that $\mathcal{H}(H+1|h)$, the probability of death in state h , is perturbed such that **Assumption 8** is no longer satisfied, so that **Theorem 5** no longer applies. The third setting increases the likelihood of receiving kidney offers to better assess the impact of retransplantation. Value iteration is used to solved MDPs.

We now describe some of the model parameters, with full details of the remaining parameter settings provided in **Appendix B**. As mentioned earlier, the patient state h is defined based on the EPTS score, a deterministic function of age (or equivalently, the decision period) in this case. If the patient is in state $h, h < H$ and chooses W , they either die or transition to state $h+1$ at the next epoch. We further assume that $\mathcal{H}(H+1|h)$ is an increasing affine function of h . Consequently, for $h < H$,

$$\mathcal{H}(h'|h) = \begin{cases} a + b(h-1) & \text{if } h' = H+1, \\ 1 - (a + b(h-1)) & \text{if } h' = h+1, \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathcal{H}(h'|H) = \begin{cases} a + b(H-1) & \text{if } h' = H+1, \\ 1 - (a + b(H-1)) & \text{if } h' = H, \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

where $a = 0.01$ in both experiments, $b = 0.007$ in Experiment 1 and 3, and 0.006 in Experiment 2. The definition of the transition function \mathcal{Q} is similar. Other parameters are the same in all the experiments, unless specified otherwise. We set the discount factor $\lambda = 0.99$. The following parameters are assigned according to OPTN (2022a). The pmf of the kidney state k , from $k=1$ to 5, is (0.0491, 0.0323, 0.1206, 0.0347, 0.7653). The pmf of the mismatch level m , from $m=1$ to 7, is (0.0492, 0.0104, 0.0192, 0.1437, 0.2806, 0.3254, 0.1414). Because patients of age 70 and older have EPTS scores greater than 20 and OPTN (2022a) doesn't distinguish patients with EPTS scores greater than 20, we assume in this section that *the probability of a transplantation failure* depends only on k and m , not on h , and denote it by $\mathcal{D}(k, m)$. The values of $\mathcal{D}(k, m)$ are summarized in **Table 1**.

The values of the post-transplant reward $r(h, k, m)$ are calculated according to Bae et al. (2019) and provided in **Equation (10)** for $m=1$ and 7, with the rest provided in **Appendix B**.

$$r(h, k, 1) = \begin{bmatrix} 12 & 11 & 10 & 8.5 \\ 11 & 11 & 9.5 & 8.2 \\ 9.9 & 9.7 & 9 & 7.9 \\ 9.6 & 9.4 & 8.8 & 7.7 \\ 9.3 & 9.1 & 8.5 & 7.6 \\ 8.9 & 8.8 & 8.3 & 7.4 \\ 8.7 & 8.5 & 8 & 7.2 \\ 8.5 & 8.4 & 7.8 & 7.1 \\ 8.3 & 8.1 & 7.7 & 6.9 \\ 8.1 & 8 & 7.5 & 6.8 \\ 8.1 & 8 & 7.5 & 6.8 \\ 8 & 7.9 & 7.4 & 6.7 \\ 7.8 & 7.7 & 7.3 & 6.6 \\ 7.7 & 7.6 & 7.2 & 6.6 \\ 7.7 & 7.6 & 7.1 & 6.5 \\ 7.6 & 7.5 & 7.1 & 6.5 \end{bmatrix}, \quad r(h, k, 7) = \begin{bmatrix} 6 & 5.9 & 5.8 & 5.5 \\ 5.9 & 5.9 & 5.8 & 5.5 \\ 5.8 & 5.8 & 5.7 & 5.4 \\ 5.8 & 5.7 & 5.6 & 5.3 \\ 5.7 & 5.7 & 5.6 & 5.3 \\ 5.6 & 5.6 & 5.5 & 5.2 \\ 5.6 & 5.6 & 5.4 & 5.1 \\ 5.5 & 5.5 & 5.3 & 5.1 \\ 5.5 & 5.4 & 5.3 & 5 \\ 5.4 & 5.4 & 5.3 & 5 \\ 5.4 & 5.4 & 5.3 & 5 \\ 5.4 & 5.4 & 5.2 & 4.9 \\ 5.4 & 5.3 & 5.2 & 4.9 \\ 5.3 & 5.3 & 5.1 & 4.8 \\ 5.3 & 5.2 & 5.1 & 4.8 \\ 5.3 & 5.2 & 5.1 & 4.8 \end{bmatrix}. \quad (10)$$

Table 1. The probability of a transplantation failure $\mathcal{D}(k, m)$.

$\mathcal{D}(k, m)$	$m = 1$	$m > 1$
$k = 1$	0.017	0.041
$k = 2$	0.037	0.061
$k = 3$	0.047	0.071
$k = 4$	0.073	0.095

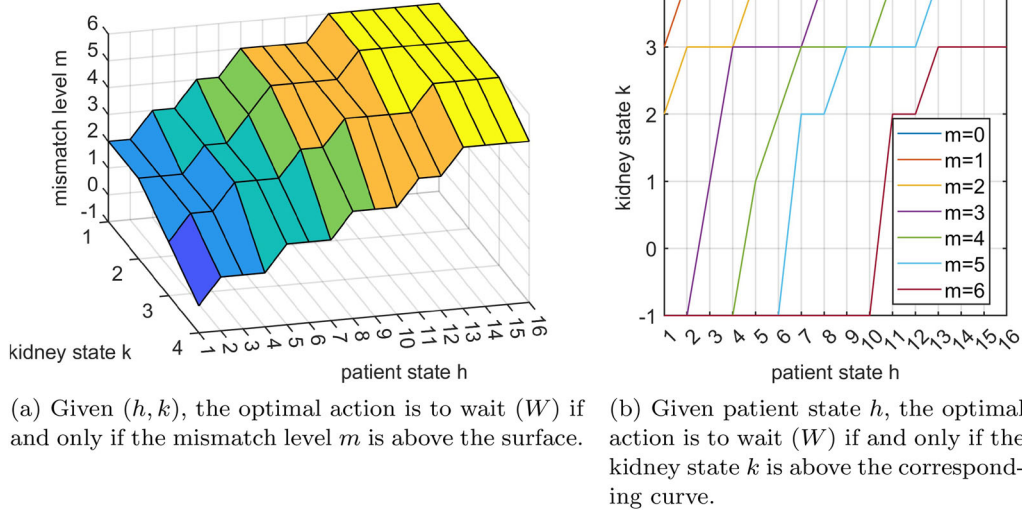


Figure 3. The optimal policy d_1 .

Experiment 1

The optimal policy d_1 is given in Figure 3. All three types of control limit optimal policies exist. For each mismatch level, we plot the projection of the optimal policy on the $h-k$ plane.

The Q-function is the expected total discounted reward starting from a given state, taking a given action, and following the optimal policy thereafter. We plot in Figure 4 the Q-function, given by Equation (5).

Note that the value of the Q-function depends only on the patient state h if the action is to wait. When the mismatch level is low, it is more beneficial to accept an offer unless it is of very low quality. As the mismatch level increases, the optimal action is to wait and then switch to transplant if the patient state h is below some control limit (i.e., the patient health is worse than some threshold). For fixed mismatch level, as the patient state h worsens, the optimal policy tends to accept low-quality kidneys (i.e., those with high KDPI) that are rejected at lower h 's. At high mismatch levels, the decision maker should choose transplant only when the patient is in severe health status (i.e., h is sufficiently large).

Next, we compare with an optimal policy derived for an MDP model that does not include the mismatch level as a state variable, i.e., the acceptance decision making depends only on the patient and kidney states, and does not explicitly model transplantation failure or retransplantation. In the case of the latter, there is a terminal reward that is the mean over the mismatch distribution. We compute its optimal policy q_1 , shown in Figure 5. Both patient-based and kidney-based control limit optimal policies exist. The control limit curve of policy q_1 lies between curves of policy d_1 for mismatch level $m = 4$ and 5.

To evaluate the policy $q_1(h, k)$ in the original MDP with a state vector (h, k, m) , we define and implement a new policy $\hat{d}_1(h, k, m) := q_1(h, k), \forall h, k, m$. i.e., $\hat{d}_1(h, k, m)$ does not account for the mismatch level m and produces the same action as $q_1(h, k)$. We compare its value function $V_{\hat{d}_1}(h, k, m)$ with the optimal one $V_{d_1}(h, k, m)$ in Figure 6. Compared with policy d_1 , when h is low (i.e., the patient is healthy), policy \hat{d}_1 rejects offers with low mismatch level, which would have been very beneficial to patients. As a result, policy d_1 significantly outperforms \hat{d}_1 when the mismatch level m is low, e.g., $m \leq 3$. At intermediate mismatch levels, the two policies behave similarly. When the mismatch level is high, e.g., $m \geq 6$, the two policies behave differently only when h is high (i.e., the patient is in poor health state), where the difference of the Q-functions (i.e., the expected overall benefit) between accepting and rejecting an offer is small, as shown in Figure 4. Thus, the performances of the two policies are close when the mismatch level is intermediate to high.

Experiment 2

The optimal policy d_2 is shown in Figure 7. The kidney-based and match-based control limit optimal policies still exist, but the patient-based optimal policy for mismatch level equal to 7 is not a control limit policy.

Experiment 3

In Experiment 1, the probability of receiving a kidney per period is below 25%, while the chance of transplantation failure is below 10%. Even if the decision maker always chooses transplantation when a kidney is available, the probability of transplantation failure

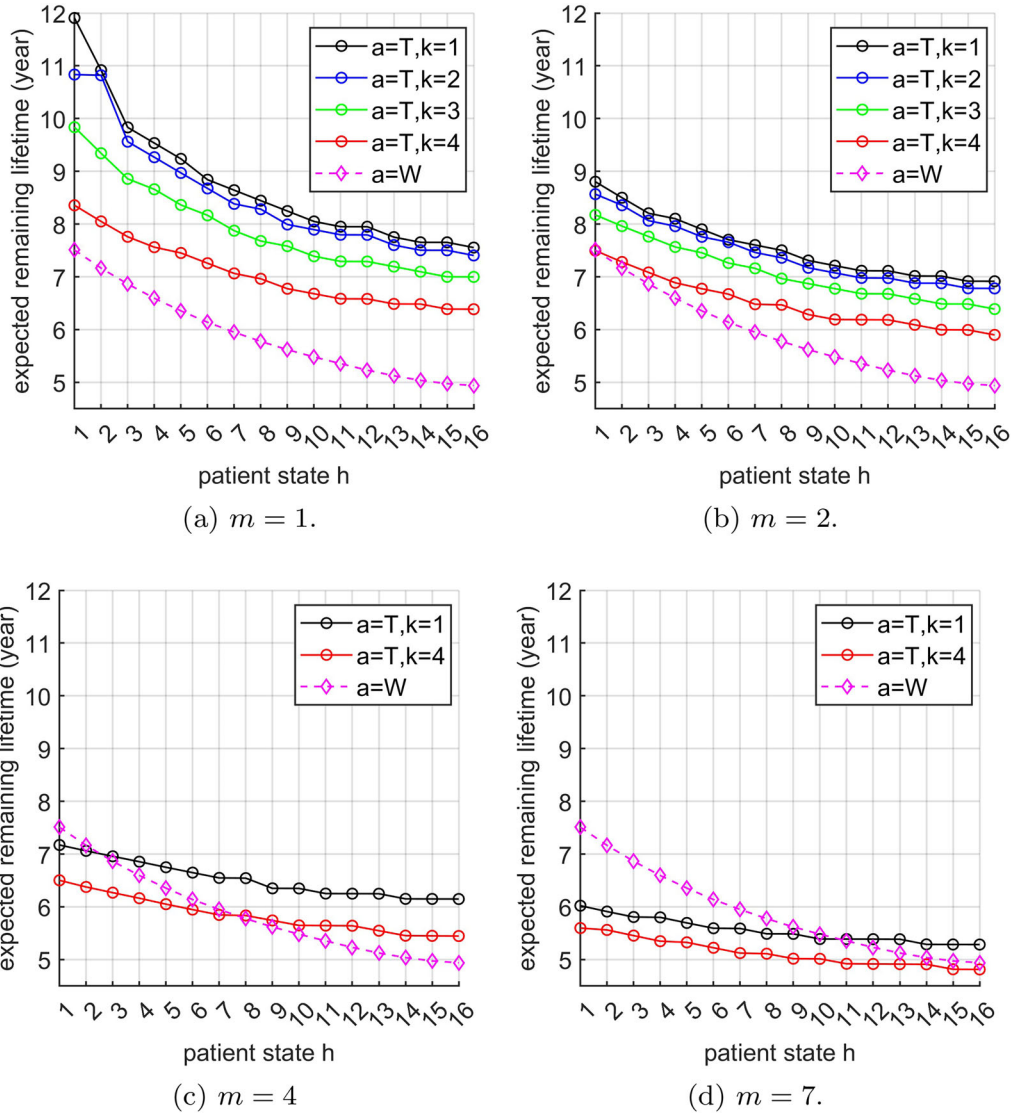


Figure 4. Q-function for different mismatch levels.

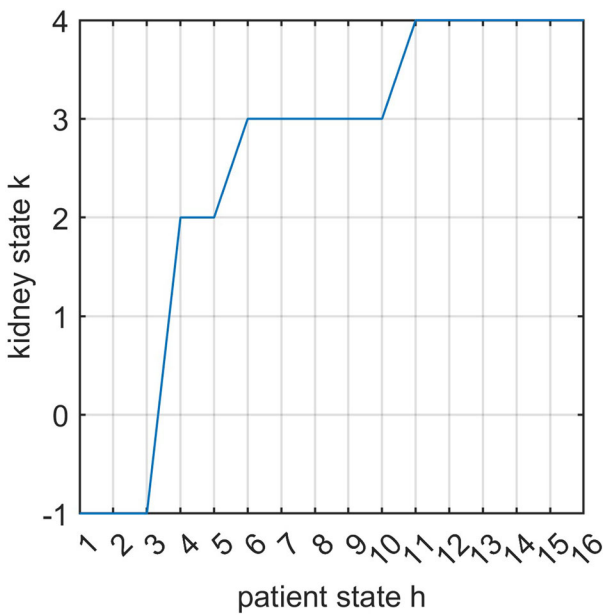


Figure 5. Policy q_1 : given patient state h , the optimal action is to wait (W) if and only if the kidney state k is above the curve.

remains below $1/40$ per period, indicating that retransplantation is rare. Notably, even assuming every transplantation is successful, the optimal policy would remain unchanged. Thus, in Experiment 1, policy d_1 outperforms \hat{d}_1 primarily because d_1 accounts for the mismatch level, while the impact of retransplantation is relatively minor.

To assess the impact of retransplantation separately from the mismatch level, we consider an MDP model where transplantation failure occurs, but the mismatch level is hidden—similar to the derivation procedure of policy q_1 which hides both factors. The transplantation reward and the probability of transplantation failure when the mismatch level is hidden are calculated by taking the expectation over the distribution of the mismatch level. To highlight the effect of transplantation failure, we double both the probability of receiving a kidney offer and the probability of failure, leaving other parameters unchanged. We compute the optimal policy q_3 and compare it with policy q_1 , as shown in Figure 8.

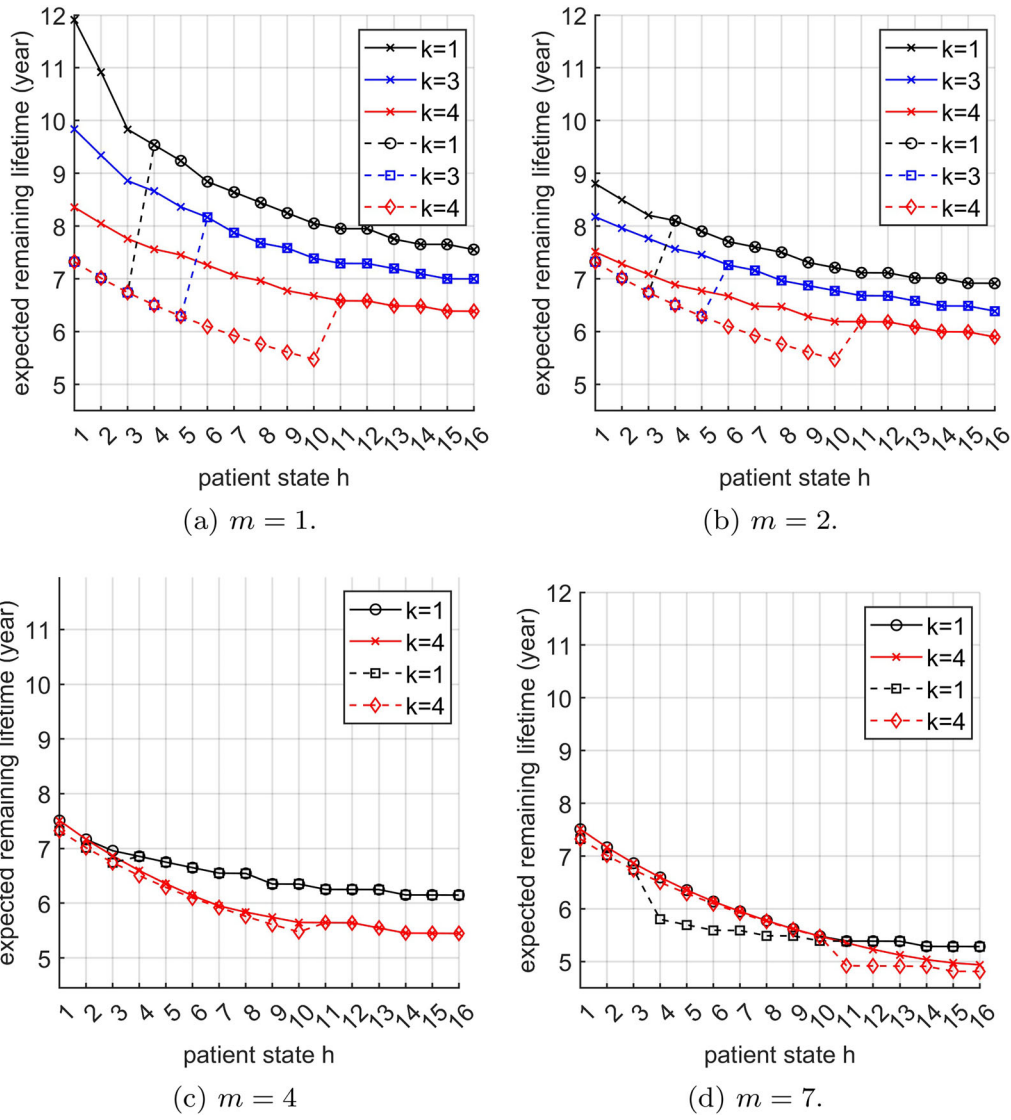


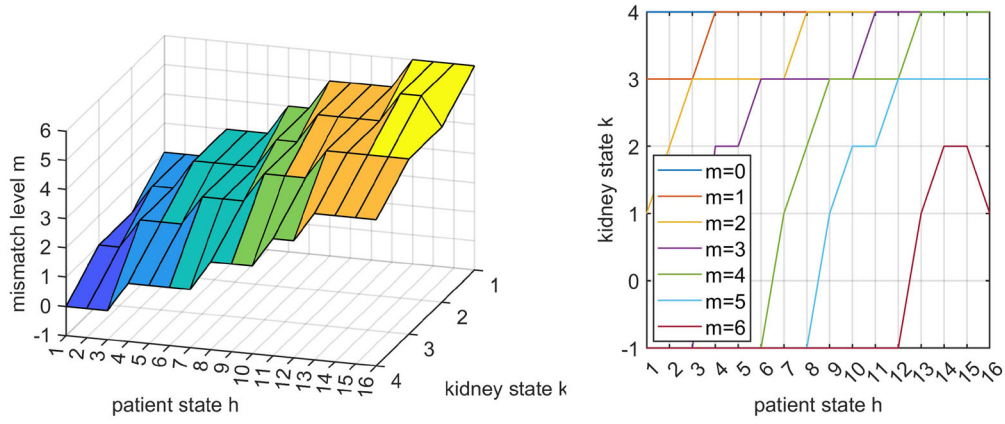
Figure 6. Value functions $V_{d_1}(h, k, m)$ and $V_{\hat{d}_1}(h, k, m)$ for different mismatch levels, with V_{d_1} represented by solid lines and $V_{\hat{d}_1}$ by dashed lines. For policy \hat{d}_1 , fixing m , curves for different values of k overlap with each other when $h \leq 3$ because \hat{d}_1 is always W for $h \leq 3$, independent of m . Since both d_1 and \hat{d}_1 are control limit policies, fixing k and m , if h is greater than control limits of both policies, both policies choose transplant. Moreover, Equation (9) implies that the patient state h_t is a nondecreasing function of time, i.e., the patient health never improves. So, fixing k and m , curves for both policies overlap when h is greater than the maximum of control limits of both (patient-based) policies, which is a function of both k and m .

The control limit curve of policy q_3 falls below that of policy q_1 for $h \leq 5$, and lies above it for $h = 12$ and 13 . This aligns with Assumption 9 and the trends in terminal transplantation rewards shown in Equation (10). Specifically, Assumption 9 suggests that transplantation failure has a more significant impact on healthier patients, while Equation (10) indicates that the transplantation reward decline rapidly as the patient state worsens when $h \leq 5$. These observations imply that the penalty for transplantation failure is more substantial for patients in better state, leading to a stricter acceptance criterion (i.e., a lower control limit curve). For $h > 6$, the transplantation reward decreases only slightly as patient state deteriorates. Specifically, for patients in intermediate or worse states, even if they experience a transplantation failure and their state worsens, the decline in transplantation reward is small. Rather

than imposing a penalty, explicitly modeling the transplantation failure event essentially provides patients in poor states with more opportunities for retransplantation, making them more inclined to accept a kidney.

5. Conclusions and future research

We consider the problem of sequentially accepting or declining possibly incompatible kidney offers by undergoing desensitization therapies through the use of an MDP model that captures the effect of both quality and compatibility by explicitly including them as state variables. We derive structural properties of the model, specifically, characterizing control limit-type policies, including patient-based, kidney-based and match-based control limit optimal policies, by providing reasonable sufficient



(a) Given (h, k) , the optimal action is to wait (W) if and only if the mismatch level m is above the surface.

(b) Given patient state h , the optimal action is to wait (W) if and only if the kidney state k is above the corresponding curve.

Figure 7. The optimal policy d_2 . In the graph on the right, for mismatch level $m = 7$, the optimal policy is not of control limit type.

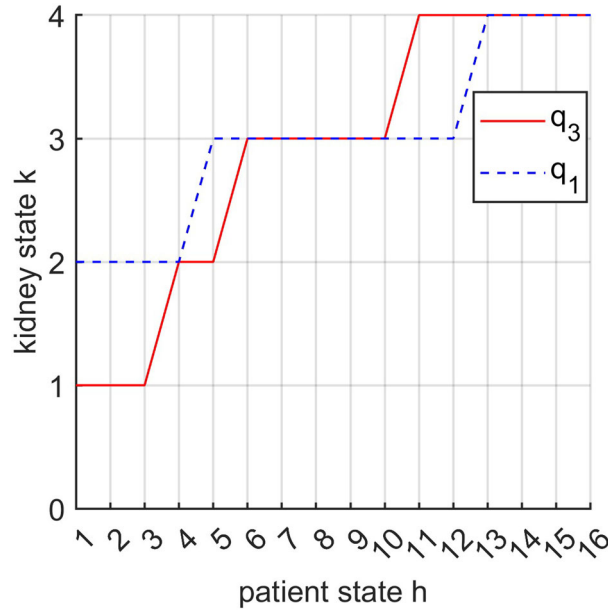


Figure 8. Policies q_3 (solid line) and q_1 (dashed line). The control limit curve of policy q_3 falls below that of policy q_1 for $h \leq 5$, and lies above it for $h = 12$ and 13 .

conditions to guarantee their existence. Numerical experiments on a stylized example based on realistic data (OPTN, 2022a) illustrate that the potential gains from taking both quality and compatibility into consideration and allowing the option of accepting a mismatched kidney can be on the order of an additional year of life expectancy for a 70-year old, whose expected remaining lifetime is less than 5 years without kidney transplantation.

Our results have concrete applications in clinical decision-making and policy design. By explicitly accounting for patient state, kidney quality, and compatibility, our model could be integrated into decision-support tools that personalize decision-making based on patient conditions, offering simple acceptance thresholds as references for surgeons.

The model's value function and Q-function quantify the gains of accepting a transplant versus waiting, providing clear measures of "good quality" and "low mismatch level." This information aids surgeons in making informed decisions and enhances transparency by enabling physicians to clearly explain trade-offs to patients, improving shared decision-making. Beyond clinical use, the acceptance criteria and quantified gains from our model could refine kidney allocation policies. These measures could enhance allocation scoring models to improve the overall transplant outcomes while balancing equity, particularly for patients with rare HLA types or blood types. The patient-specific acceptance thresholds could guide kidney match policies by estimating the likelihood of acceptance, reducing organ rejection

and wastage, and ensuring more effective kidney utilization.

Our model uses aggregated scores like EPTS and KDPI to model the state of patients and donor kidneys, making it feasible to compute the optimal policy and providing a straightforward reference tool for deciding whether to accept or decline marginal kidney offers. However, transplant surgeons consider a broader range of factors, including demographic details and additional lab results not covered by EPTS and KDPI. These factors significantly influence transplant outcomes and patient priority rankings in KAS or KPD, affecting kidney offer availability. Additionally, these features may have complex correlations that aggregated scores cannot fully capture. While an MDP model with a state vector incorporating these features would be more accurate and realistic, it would also increase the dimension of the state space, making computation intractable. Exploring the impact of these additional factors and balancing computational feasibility with model accuracy is a promising area for future research. Based on our research, it is evident that the patient's decision can significantly differ depending on the level of mismatch. Therefore, from the standpoint of policy makers, integrating our model—which explicitly accounts for the possibility of accepting an incompatible kidney—into the kidney allocation problem could potentially help further reduce organ shortages and enhance overall social welfare.

Acknowledgments

The authors thank Naoru Koizumi for many useful discussions on incompatible kidney transplantation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported in part by the National Science Foundation under Grant IIS-2123684 and by AFOSR under Grant FA95502010211.

References

- Ahn, J.-H., & Hornberger, J. C. (1996). Involving patients in the cadaveric kidney transplant allocation process: A decision-theoretic perspective. *Management Science*, 42(5), 629–641. <https://doi.org/10.1287/mnsc.42.5.629>
- Aimaretti, L. A., & Arze, S. (2016). Preemptive renal transplantation—The best treatment option for terminal chronic renal failure. *Transplantation Proceedings*, 48, 609–611. <https://doi.org/10.1016/j.transproceed.2016.02.047>
- Akan, M., Alagoz, O., Ata, B., Erenay, F. S., & Said, A. (2012). A broader view of designing the liver allocation system. *Operations Research*, 60(4), 757–770. <https://doi.org/10.1287/opre.1120.1064>
- Alagoz, O., Hsu, H., Schaefer, A. J., & Roberts, M. S. (2010). Markov decision processes: A tool for sequential decision making under uncertainty. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, 30(4), 474–483. <https://doi.org/10.1177/0272989X09353194>
- Alagoz, O., Maillart, L. M., Schaefer, A. J., & Roberts, M. S. (2004). The optimal timing of living-donor liver transplantation. *Management Science*, 50(10), 1420–1430. <https://doi.org/10.1287/mnsc.1040.0287>
- Alagoz, O., Maillart, L. M., Schaefer, A. J., & Roberts, M. S. (2007a). Choosing among living-donor and cadaveric livers. *Management Science*, 53(11), 1702–1715. <https://doi.org/10.1287/mnsc.1070.0726>
- Alagoz, O., Maillart, L. M., Schaefer, A. J., & Roberts, M. S. (2007b). Determining the acceptance of cadaveric livers using an implicit model of the waiting list. *Operations Research*, 55(1), 24–36. <https://doi.org/10.1287/opre.1060.0329>
- Ashlagi, I., & Roth, A. E. (2021). Kidney exchange: An operations perspective. *Management Science*, 67(9), 5455–5478. <https://doi.org/10.1287/mnsc.2020.3954>
- Ata, B., Skaro, A., & Tayur, S. (2017). OrganJet: Overcoming geographical disparities in access to deceased donor kidneys in the United States. *Management Science*, 63(9), 2776–2794. <https://doi.org/10.1287/mnsc.2016.2487>
- Avila-Godoy, G., & Fernández-Gauchera, E. (1998). Controlled Markov chains with exponential risk-sensitive criteria: Modularity, structured policies and applications. In *Proceedings of the 37th IEEE Conference on Decision and Control (Cat. No. 98CH36171)*, Volume 1, pp. 778–783. IEEE. <https://doi.org/10.1109/CDC.1998.760782>
- Bae, S., Massie, A. B., Thomas, A. G., Bahn, G., Luo, X., Jackson, K. R., Ottmann, S. E., Brennan, D. C., Desai, N. M., Coresh, J., Segev, D. L., & Wang, G. J. (2019). Who can tolerate a marginal kidney? Predicting survival after deceased donor kidney transplant by donor-recipient combination. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, 19(2), 425–433. <https://doi.org/10.1111/ajt.14978>
- Batun, S., Schaefer, A. J., Bhandari, A., & Roberts, M. S. (2018). Optimal liver acceptance for risk-sensitive patients. *Service Science*, 10(3), 320–333. <https://doi.org/10.1287/serv.2018.0215>
- Bendersky, M., & David, I. (2016). Deciding kidney-offer admissibility dependent on patients' lifetime failure rate. *European Journal of Operational Research*, 251(2), 686–693. <https://doi.org/10.1016/j.ejor.2015.12.009>
- Bertsekas, D. P. (2020). *Dynamic programming and optimal control* (4th ed., Vol.I). Athena Scientific.
- Clayton, P. A., & Coates, P. T. (2017). Are sensitized patients better off with a desensitization transplant or waiting on dialysis? *Kidney International*, 91(6), 1266–1268. <https://doi.org/10.1016/j.kint.2017.04.004>
- C-STAR (2013). KT candidacy calculator for patients 65+. <http://www.transplantmodels.com/candidate65/>
- C-STAR (2015). Live donor KDPI calculator. <http://www.transplantmodels.com/lkdpi/>

- C-STAR (2018). KDPI-EPTS survival benefit estimator. <http://www.transplantmodels.com/kdpi-epts/>
- David, I., & Yechiali, U. (1985). A time-dependent stopping problem with application to live organ transplants. *Operations Research*, 33(3), 491–504. <https://doi.org/10.1287/opre.33.3.491>
- Fan, W., Zong, Y., & Kumar, S. (2022). Optimal treatment of chronic kidney disease with uncertainty in obtaining a transplantable kidney: An MDP-based approach. *Annals of Operations Research*, 316(1), 269–302. <https://doi.org/10.1007/s10479-020-03779-2>
- Fiorentino, M., Gallo, P., Giliberti, M., Colucci, V., Schena, A., Stallone, G., Gesualdo, L., & Castellano, G. (2021). Management of patients with a failed kidney transplant: What should we do? *Clinical Kidney Journal*, 14(1), 98–106. <https://doi.org/10.1093/ckj/sfaa094>
- Flores-Hernández, R. M., & Montes-de Oca, R. (2007). Monotonicity of minimizers in optimization problems with applications to Markov control processes. *Kybernetika*, 43(3), 347–368.
- Grams, M. E., Kucirka, L. M., Hanrahan, C. F., Montgomery, R. A., Massie, A. B., & Segev, D. L. (2012). Candidacy for kidney transplantation of older adults. *Journal of the American Geriatrics Society*, 60(1), 1–7. <https://doi.org/10.1111/j.1532-5415.2011.03652.x>
- Husain, S. A., King, K. L., Pastan, S., Patzer, R. E., Cohen, D. J., Radhakrishnan, J., & Mohan, S. (2019). Association between declined offers of deceased donor kidney allograft and outcomes in kidney transplant candidates. *JAMA Network Open*, 2(8), e1910312–e1910312. <https://doi.org/10.1001/jamanetworkopen.2019.10312>
- Kaufman, D., Schaefer, A. J., & Roberts, M. S. (2017). Living-donor liver transplantation timing under ambiguous health state transition probabilities. *SSRN* 3003590.
- Kim, J. J., Fuggle, S. V., & Marks, S. D. (2021). Does HLA matching matter in the modern era of renal transplantation? *Pediatric Nephrology (Berlin, Germany)*, 36(1), 31–40. <https://doi.org/10.1007/s00467-019-04393-6>
- King, K. L., Husain, S. A., Cohen, D. J., Schold, J. D., & Mohan, S. (2022). The role of bypass filters in deceased donor kidney allocation in the United States. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, 22(6), 1593–1602. <https://doi.org/10.1111/ajt.16967>
- Ko, E. J., Yu, J. H., Yang, C. W., Chung, B. H., & Group, K. O. T. R. S., Korean Organ Transplantation Registry Study Group (2017). Clinical outcomes of ABO- and HLA-incompatible kidney transplantation: A nationwide cohort study. *Transplant International: Official Journal of the European Society for Organ Transplantation*, 30(12), 1215–1225. <https://doi.org/10.1111/tri.12979>
- Konvalinka, A., & Tinckam, K. (2015). Utility of HLA antibody testing in kidney transplantation. *Journal of the American Society of Nephrology: JASN*, 26(7), 1489–1502. <https://doi.org/10.1681/ASN.2014080837>
- Koo, T. Y., & Yang, J. (2015). Current progress in ABO-incompatible kidney transplantation. *Kidney Research and Clinical Practice*, 34(3), 170–179. <https://doi.org/10.1016/j.krcp.2015.08.005>
- Kuppachi, S., & Axelrod, D. A. (2020). Desensitization strategies: Is it worth it? *Transplant International: Official Journal of the European Society for Organ Transplantation*, 33(3), 251–259. <https://doi.org/10.1111/tri.13532>
- Lentine, K. L., Smith, J. M., Miller, J. M., Bradbrook, K., Larkin, L., Weiss, S., Handarova, D. K., Temple, K., Israni, A. K., & Snyder, J. J. (2023). OPTN/SRTR 2021 annual data report: Kidney. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, 23(2 Suppl 1), S21–S120. <https://doi.org/10.1016/j.ajt.2023.02.004>
- Lovejoy, W. S. (1987). Some monotonicity results for partially observed Markov decision processes. *Operations Research*, 35(5), 736–743. <https://doi.org/10.1287/opre.35.5.736>
- Mahawar, K. K., & Bal, A. M. (2004). Role of HLA matching in liver transplant. *Transplantation*, 78(2), 643. <https://doi.org/10.1097/00007890-200407271-01730>
- Massie, A. B., Leanza, J., Fahmy, L. M., Chow, E. K., Desai, N. M., Luo, X., King, E. A., Bowring, M. G., & Segev, D. L. (2016). A risk index for living donor kidney transplantation. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, 16(7), 2077–2084. <https://doi.org/10.1111/ajt.13709>
- Miehling, E., & Teneketzis, D. (2020). Monotonicity properties for two-action partially observable Markov decision processes on partially ordered spaces. *European Journal of Operational Research*, 282(3), 936–944. <https://doi.org/10.1016/j.ejor.2019.10.003>
- Morath, C., Zeier, M., Döhler, B., Opelz, G., & Süsal, C. (2017). ABO-incompatible kidney transplantation. *Frontiers in Immunology*, 8, 234. <https://doi.org/10.3389/fimmu.2017.00234>
- Oh, S., & Özer, Ö. (2016). Characterizing the structure of optimal stopping policies. *Production and Operations Management*, 25(11), 1820–1838. <https://doi.org/10.1111/poms.12579>
- Opelz, G., & Döhler, B. (2007). Effect of human leukocyte antigen compatibility on kidney graft survival: Comparative analysis of two decades. *Transplantation*, 84(2), 137–143. <https://doi.org/10.1097/01.tp.0000269725.74189.b9>
- OPTN (2022a). Eliminate use of DSA and region from kidney allocation: One year post-implementation monitoring report. https://optn.transplant.hrsa.gov/media/p2oc3ada/data_report_kidney_full_20220624_1.pdf
- OPTN (2022b). EPTS calculator. <https://optn.transplant.hrsa.gov/data/allocation-calculators/epts-calculator/>
- OPTN (2022c). Guide to calculating & interpreting EPTS. https://optn.transplant.hrsa.gov/media/1511/guide_to_calculating_interpreting_epts.pdf
- OPTN (2022d). Guide to calculating & interpreting KDPI. https://optn.transplant.hrsa.gov/media/1512/guide_to_calculating_interpreting_kdpi.pdf
- OPTN (2022e). View data source. <https://optn.transplant.hrsa.gov/data/view-data-reports/>
- OPTN (2023). Organ procurement and transplantation network (OPTN) policies. https://optn.transplant.hrsa.gov/media/eavh5bf3/optn_policies.pdf
- Puterman, M. L. (2014). *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons.
- Ren, X., Fu, M. C., & Marcus, S. I. (2023a). Sensitivity analysis for stopping criteria with application to organ transplantations. In *Proceedings of the 2023 Winter*

- Simulation Conference (WSC)*, pp. 504–515. IEEE. <https://doi.org/10.1109/WSC60868.2023.10408603>
- Ren, X., Fu, M. C., & Marcus, S. I. (2023b). Stochastic control for organ donations: A review. *Systems & Control Letters*, 173, 105476. <https://doi.org/10.1016/j.sysconle.2023.105476>
- Ross, S. M. (1996). *Stochastic processes* (2nd ed.). Wiley New York.
- Rydberg, L., Skogsberg, U., & Mölne, J. (2007). ABO antigen expression in graft tissue: Is titration against donor erythrocytes relevant? *Transplantation*, 84(12 Suppl), S10–S12. <https://doi.org/10.1097/01.tp.0000296017.44394.e2>
- Serfozo, R. F. (1976). Monotone optimal policies for Markov decision processes. *Stochastic Systems: Modeling, Identification and Optimization, II*, 202–215.
- Smith, J. E., & McCardle, K. F. (2002). Structural properties of stochastic dynamic programs. *Operations Research*, 50(5), 796–809. <https://doi.org/10.1287/opre.50.5.796.365>
- SRTR (2023). Kidney transplant decision aid. <https://www.srtr.org/tools/kidney-transplant-decision-aid/>
- Su, X., & Zenios, S. (2004). Patient choice in kidney allocation: The role of the queueing discipline. *Manufacturing & Service Operations Management*, 6(4), 280–301. <https://doi.org/10.1287/msom.1040.0056>
- Su, X., & Zenios, S. A. (2005). Patient choice in kidney allocation: A sequential stochastic assignment model. *Operations Research*, 53(3), 443–455. <https://doi.org/10.1287/opre.1040.0180>
- Su, X., & Zenios, S. A. (2006). Recipient choice can address the efficiency-equity trade-off in kidney transplantation: A mechanism design model. *Management Science*, 52(11), 1647–1660. <https://doi.org/10.1287/mnsc.1060.0541>
- Tunç, S., Sandıkçı, B., & Tanrıöver, B. (2022). A simple incentive mechanism to alleviate the burden of organ wastage in transplantation. *Management Science*, 68(8), 5980–6002. <https://doi.org/10.1287/mnsc.2021.4203>

Appendix A. Proofs

Define

$$\begin{aligned} U(h, k) &:= \sum_{m \in S_M} \mathcal{M}(m) V(h, k, m), h \in S_H, k \in S_K, \\ W(h, m) &:= \sum_{k \in S_K} \mathcal{K}(k|h) V(h, k, m), h \in S_H, m \in S_M, \end{aligned} \quad (\text{A1})$$

which will be used in the proofs. Note that $U(h, k)$ and $W(h, m)$ have similar interpretations as $v(h)$, which can be interpreted as the expected total discounted reward when the patient state is h . Correspondingly, we define the following quantities associated with each iteration of the value iteration algorithm:

$$\begin{aligned} U_n(h, k) &= \sum_{m \in S_M} \mathcal{M}(m) V_n(h, k, m), h \in S_H, k \in S_K, \\ W_n(h, m) &= \sum_{k \in S_K} \mathcal{K}(k|h) V_n(h, k, m), h \in S_H, m \in S_M. \end{aligned} \quad (\text{A2})$$

We begin with [Lemmas 1](#) and [2](#), which are useful for proving [Lemma 3](#), an important intermediate result for proving [Theorem 2](#).

Lemma 1. (Puterman, 2014) Let $\{x_j\}_{j \in \mathbb{N}}, \{x'_j\}_{j \in \mathbb{N}}$ be real-valued nonnegative sequences satisfying

$$\sum_{j=k}^{\infty} x_j \geq \sum_{j=k}^{\infty} x'_j$$

for all k , with equality holding for $k = 0$. Suppose $v_{j+1} \geq v_j, \forall j$, then

$$\sum_{j=0}^{\infty} v_j x_j \geq \sum_{j=0}^{\infty} v_j x'_j.$$

[Lemma 2](#) immediately follows from [Lemma 1](#).

Lemma 2. If \mathcal{H} is stochastically increasing and $f: \mathbb{R} \mapsto \mathbb{R}$ is nondecreasing, then

$$\sum_{h \in S_H} f(h) \mathcal{H}(h|h_1) \geq \sum_{h \in S_H} f(h) \mathcal{H}(h|h_2)$$

for any $h_1 \geq h_2$.

Lemma 3. If [Assumptions 1](#) through [6](#) hold, and if $V(h, k, m)$ is nonincreasing in h and k , then $v(h)$ is nonincreasing in h .

Proof of Lemma 3. Recall the Bellman equation [Theorem 1](#) where the second equation follows by noticing that rejecting an offer is equivalent to the offer being unavailable, and that $A^*(h, k, m)$ is the set of optimal actions at state (h, k, m) .

Fix $h \leq H$ and $m \in S_M$. Define

$$K_T := \{k \in S_K | T \in A^*(h+1, k, m)\}.$$

Note that if $k \in K_T$, then $k \leq K$, i.e., the kidney offer is available. Then, [Assumption 6](#) implies that $\mathcal{K}(k|h+1) \leq \mathcal{K}(k|h)$.

For $k \in K_T$, since $V(h, k, m)$ is nonincreasing in h and k ,

$$\begin{aligned} &V(h, k, m) \mathcal{K}(k|h) - V(h+1, k, m) \mathcal{K}(k|h+1) \\ &\geq V(h+1, k, m) (\mathcal{K}(k|h) - \mathcal{K}(k|h+1)) \\ &\geq V(h+1, K+1, m) (\mathcal{K}(k|h) - \mathcal{K}(k|h+1)). \end{aligned}$$

For $k \notin K_T$, $V(h+1, k, m) = V(h+1, K+1, m)$. Hence,

$$\begin{aligned} &V(h, k, m) \mathcal{K}(k|h) - V(h+1, k, m) \mathcal{K}(k|h+1) \\ &\geq V(h, K+1, m) \mathcal{K}(k|h) - V(h+1, K+1, m) \mathcal{K}(k|h+1) \\ &\geq V(h+1, K+1, m) (\mathcal{K}(k|h) - \mathcal{K}(k|h+1)). \end{aligned}$$

Therefore,

$$\begin{aligned} W(h, m) - W(h+1, m) &= \sum_{k \in K_T} (V(h, k, m) \mathcal{K}(k|h) - V(h+1, k, m) \mathcal{K}(k|h+1)) \\ &\quad + \sum_{k \notin K_T} (V(h, k, m) \mathcal{K}(k|h) - V(h+1, k, m) \mathcal{K}(k|h+1)) \\ &\geq \sum_{k \in S_K} (\mathcal{K}(k|h) - \mathcal{K}(k|h+1)) V(h+1, K+1, m) \\ &= 0, \end{aligned}$$

because $\sum_{k \in S_K} (\mathcal{K}(k|h) - \mathcal{K}(k|h+1)) = 0$. It follows that

$$\begin{aligned} v(h) - v(h+1) &= \sum_{m \in S_M} \mathcal{M}(m) (W(h, m) - W(h+1, m)) \\ &\geq 0. \end{aligned} \quad \blacksquare$$

Proof of Theorem 2. Prove by induction. Consider the value iteration algorithm [Equation \(2\)](#) starting at $V_0(h, k, m) = 0, \forall h, k$, and m . Outline of the proof: at each step n , we show:

1. $V_n(h, k, m)$ is nonincreasing in k . Define

$$\begin{aligned} K_n^*(h, m) &= \max\{k \in S_K \mid V_n(h, k, m) \\ &= (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)(c(h) \\ &+ \lambda \sum_{h' \in S_H} v_{n-1}(h') \mathcal{Q}(h'|h))\}, \end{aligned}$$

where we assume without loss of generality (as we show below) that

$$\begin{aligned} \{k \in S_K \mid V_n(h, k, m) &= (1 - \mathcal{D}(h, k, m))r(h, k, m) \\ &+ \mathcal{D}(h, k, m)(c(h) \\ &+ \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h))\} \neq \emptyset. \end{aligned} \quad (\text{A3})$$

We show that for fixed h and m , if $k \leq K_n^*(h, m)$,

$$\begin{aligned} V_n(h, k, m) &= (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)(c(h) \\ &+ \lambda \sum_{h' \in S_H} v_{n-1}(h') \mathcal{Q}(h'|h)), \end{aligned}$$

and $V_n(h, k, m) = c(h) + \lambda \sum_{h' \in S_H} v_{n-1}(h') \mathcal{H}(h'|h)$ if $k > K_n^*(h, m)$. Then, we show that $V_n(h, k, m)$ is nonincreasing in k on both pieces. Note that the proof of $V_n(h, k, m)$ being nonincreasing is trivial if Equation (A3) does not hold, because $V_n(h, k, m) = c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h)$, $\forall k \in S_K$, which does not depend on k . From now on, we use the abbreviation K_n^* for $K_n^*(h, m)$.

2. $V_n(h, k, m)$ and $v_n(h)$ are nonincreasing in h , where we will use Lemma 3.

Finally, the value function V , as the limit of the sequence $\{V_n\}$, is nonincreasing in both h and k .

Initial step: by Equation (2), for any h and m , we have

$$\begin{aligned} V_1(h, k, m) &= \begin{cases} \max((1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h), c(h)) & \text{if } k = 0, \dots, K, \\ c(h) & \text{if } k = K + 1. \end{cases} \end{aligned}$$

Then,

$$\begin{aligned} (1 - \mathcal{D}(h, K_1^*, m))r(h, K_1^*, m) + \mathcal{D}(h, K_1^*, m)c(h) &\geq c(h), \\ \Rightarrow (1 - \mathcal{D}(h, K_1^*, m))(r(h, K_1^*, m) - c(h)) &\geq 0, \\ \Rightarrow r(h, K_1^*, m) - c(h) &\geq 0, \\ \Rightarrow r(h, k, m) - c(h) &\geq 0, \forall k \leq K_1^*, \end{aligned}$$

where the last step follows from Assumption 1 that $r(h, k, m)$ is nonincreasing in k . Therefore, $(1 - \mathcal{D}(h, k, m))(r(h, k, m) - c(h)) \geq 0, \forall k \leq K_1^*$, i.e.,

$$(1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h) \geq c(h), k \leq K_1^*.$$

Thus,

$$V_1(h, k, m) = \begin{cases} (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h) & \text{if } k \leq K_1^*, \\ c(h) & \text{if } k > K_1^*. \end{cases}$$

For $k \leq K_1^*$, $V_1(h, k, m) = (1 - \mathcal{D}(h, k, m))(r(h, k, m) - c(h)) + c(h)$. Since both $(1 - \mathcal{D}(h, k, m))$ and $(r(h, k, m) - c(h))$ are positive and nonincreasing in k for $k \leq K_1^*$, $V_1(h, k, m)$ is nonincreasing in k for $k \leq K_1^*$. Moreover, $V_1(h, k, m) \geq c(h)$ for $k \leq K_1^*$. Therefore, $V_1(h, k, m)$ is nonincreasing in k .

From Equation (2), we see that $V_1(h, k, m)$ takes the value of either term in the maximization. Therefore, to show $V_1(h, k, m) \geq V_1(h + 1, k, m)$ for any k and m , it suffices to consider the following four cases:

1. $V_1(h, k, m) = c(h), V_1(h + 1, k, m) = c(h + 1)$.
Proving $V_1(h, k, m) \geq V_1(h + 1, k, m)$ is trivial, as $c(h) \geq c(h + 1)$ by Assumption 2.
2. $V_1(h, k, m) = (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h), V_1(h + 1, k, m) = c(h + 1)$. Then,
$$\begin{aligned} V_1(h, k, m) &= (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h) \\ &\geq c(h) \geq c(h + 1) \\ &= V_1(h + 1, k, m). \end{aligned}$$
3. $V_1(h, k, m) = (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h), V_1(h + 1, k, m) = (1 - \mathcal{D}(h + 1, k, m))r(h + 1, k, m) + \mathcal{D}(h + 1, k, m)c(h + 1)$. Note that
$$(1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h) \geq c(h).$$

Then,

$$(1 - \mathcal{D}(h, k, m))(r(h, k, m) - c(h)) \geq 0 \Rightarrow r(h, k, m) - c(h) \geq 0.$$

Therefore,

$$\begin{aligned} V_1(h, k, m) - V_1(h + 1, k, m) &= (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h) - (1 - \mathcal{D}(h + 1, k, m))r(h + 1, k, m) \\ &\quad - \mathcal{D}(h + 1, k, m)c(h + 1) \\ &= (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h) - (1 - \mathcal{D}(h + 1, k, m))r(h + 1, k, m) \\ &\quad - \mathcal{D}(h, k, m)c(h + 1) + (\mathcal{D}(h, k, m) - \mathcal{D}(h + 1, k, m))c(h + 1) \\ &\geq (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h) - (1 - \mathcal{D}(h + 1, k, m))r(h + 1, k, m) \\ &\quad - \mathcal{D}(h, k, m)c(h + 1) + (\mathcal{D}(h, k, m) - \mathcal{D}(h + 1, k, m))r(h, k, m) \\ &= (1 - \mathcal{D}(h + 1, k, m))(r(h, k, m) - r(h + 1, k, m)) + \mathcal{D}(h, k, m)(c(h) - c(h + 1)) \\ &\geq 0, \end{aligned} \quad (\text{A4})$$

where the first inequality follows from the fact that $\mathcal{D}(h, k, m) \leq \mathcal{D}(h + 1, k, m)$ and $r(h, k, m) \geq c(h) \geq c(h + 1)$, and the last inequality follows from Assumptions 1 and 2.

4. $V_1(h, k, m) = c(h), V_1(h + 1, k, m) = (1 - \mathcal{D}(h + 1, k, m))r(h + 1, k, m) + \mathcal{D}(h + 1, k, m)c(h + 1)$. Note that
$$\begin{aligned} (1 - \mathcal{D}(h + 1, k, m))r(h + 1, k, m) \\ + \mathcal{D}(h + 1, k, m)c(h + 1) \\ \geq c(h + 1). \end{aligned}$$

Then,

$$\begin{aligned} (1 - \mathcal{D}(h + 1, k, m))(r(h + 1, k, m) - c(h + 1)) &\geq 0 \\ \Rightarrow r(h + 1, k, m) - c(h + 1) &\geq 0 \\ \Rightarrow c(h + 1) &\leq r(h, k, m). \end{aligned}$$

Notice that $V_1(h, k, m) \geq (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h)$. Using the same argument as Equation (A4), we have $V_1(h, k, m) \geq V_1(h + 1, k, m)$.

Thus, $V_1(h, k, m)$ is nonincreasing in h . Lemma 3 implies that $v_1(h)$ is nonincreasing in h .

Induction step: suppose that $V_n(h, k, m)$ is nonincreasing in both h and k , and $v_n(h)$ is nonincreasing in h . By Equation (2),

$$\begin{aligned}
& V_{n+1}(h, k, m) \\
&= \max \left(\begin{aligned} & (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)(c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h)) \\ & c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h) \end{aligned} \right), \\
& V_{n+1}(h, k+1, m) = c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h),
\end{aligned}$$

$\forall h \in S_H, m \in S_M, k < K+1$. Then, for any h and m ,

$$\begin{aligned}
& (1 - \mathcal{D}(h, K_n^*, m))r(h, K_n^*, m) + \mathcal{D}(h, K_n^*, m)(c(h) \\
& + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h)) \\
&= (1 - \mathcal{D}(h, K_n^*, m)) \left(r(h, K_n^*, m) - c(h) - \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h) \right) \\
&+ c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h) \\
&\geq c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h).
\end{aligned}$$

It follows that

$$\begin{aligned}
& (1 - \mathcal{D}(h, K_n^*, m)) \left(r(h, K_n^*, m) - c(h) - \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h) \right) \\
&\geq \lambda \sum_{h' \in S_H} v_n(h') (\mathcal{H}(h'|h) - \mathcal{Q}(h'|h)) \\
&\geq 0,
\end{aligned}$$

where the last inequality follows from [Assumption 5](#) that $\mathcal{Q} \geq_{st} \mathcal{H}$ and [Lemma 2](#). Since $r(h, k, m)$ is nonincreasing in k ,

$$r(h, k, m) - c(h) - \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h) \geq 0, \forall k \leq K_n^*.$$

Since $\mathcal{D}(h, k, m)$ is nondecreasing in k , $\forall k \leq K_n^*$,

$$\begin{aligned}
& (1 - \mathcal{D}(h, k, m)) \left(r(h, k, m) - c(h) - \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h) \right) \\
&\geq \lambda \sum_{h' \in S_H} v_n(h') (\mathcal{H}(h'|h) - \mathcal{Q}(h'|h)),
\end{aligned}$$

i.e.,

$$\begin{aligned}
& (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)(c(h) \\
& + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h)) \\
&\geq c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h).
\end{aligned}$$

Therefore,

$$\begin{cases} V_{n+1}(h, k, m) = (1 - \mathcal{D}(h, k, m))r(h, k, m) \\ + \mathcal{D}(h, k, m)(c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h)) & \text{if } k \leq K_n^*, \\ V_{n+1}(h, k, m) = c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h) & \text{if } k > K_n^*. \end{cases}$$

Moreover, when $k \leq K_n^*$,

$$\begin{aligned}
& (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)(c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h)) \\
&= (1 - \mathcal{D}(h, k, m)) \left(r(h, k, m) - c(h) - \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h) \right) \\
&+ c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h),
\end{aligned}$$

where both $(1 - \mathcal{D}(h, k, m))$ and $(r(h, k, m) - c(h) - \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h))$ are positive and nonincreasing in k , so $(1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)(c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h))$ is nonincreasing in k for $k \leq K_n^*$. Moreover, $(1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)(c(h) + \lambda \sum_{h' \in S_H} v_n(h') \times \mathcal{Q}(h'|h)) \geq c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h)$ for $k \leq K_n^*$. Therefore, $V_{n+1}(h, k, m)$ is nonincreasing in k .

To show $V_{n+1}(h, k, m) \geq V_{n+1}(h+1, k, m)$ for any k and m , as above, it suffices to consider the following four cases:

1. $V_{n+1}(h, k, m) = c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h)$, $V_{n+1}(h+1, k, m) = c(h+1) + \lambda \sum_{h' \in S_H} v_n(h') \times \mathcal{H}(h'|h+1)$. By [Assumption 2](#), $c(h) \geq c(h+1)$. By [Assumption 4](#), [Lemma 2](#) and $v_n(h)$ being nonincreasing, $\sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h) \geq \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h+1)$. So, $V_{n+1}(h, k, m) \geq V_{n+1}(h+1, k, m)$.
2. $V_{n+1}(h, k, m) = (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)(c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h))$, $V_{n+1}(h+1, k, m) = c(h+1) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h+1)$. Then,

$$\begin{aligned}
V_{n+1}(h, k, m) &\geq c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h) \\
&\geq c(h+1) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h+1) \\
&= V_{n+1}(h+1, k, m),
\end{aligned}$$

where the second inequality follows from the same argument as the first case.

3. $V_{n+1}(h, k, m) = (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)(c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h))$, $V_{n+1}(h+1, k, m) = (1 - \mathcal{D}(h+1, k, m))r(h+1, k, m) + \mathcal{D}(h+1, k, m)(c(h+1) + \lambda \sum_{h' \in S_H} v_n(h') \times \mathcal{Q}(h'|h+1))$. Note that $\mathbb{E}g((h, k, m), T) = (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h)$, $\forall h, k$, and m . Then,

$$\begin{aligned}
& V_{n+1}(h, k, m) - V_{n+1}(h+1, k, m) \\
&= \mathbb{E}g((h, k, m), T) + \mathcal{D}(h, k, m) \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h) \\
&- \mathbb{E}g((h+1, k, m), T) - \mathcal{D}(h+1, k, m) \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h+1) \\
&= \mathbb{E}g((h, k, m), T) - \mathbb{E}g((h+1, k, m), T) \\
&+ (\mathcal{D}(h, k, m) - \mathcal{D}(h+1, k, m)) \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h) \\
&+ \mathcal{D}(h+1, k, m) \lambda \sum_{h' \in S_H} v_n(h') (\mathcal{Q}(h'|h) - \mathcal{Q}(h'|h+1)) \\
&\geq \mathbb{E}g((h, k, m), T) - \mathbb{E}g((h+1, k, m), T) \\
&+ (\mathcal{D}(h, k, m) - \mathcal{D}(h+1, k, m)) \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h),
\end{aligned} \tag{A5}$$

where the last inequality follows from [Lemma 2](#), [Assumption 4](#) and $v_n(h)$ being nonincreasing. Note that $V_{n+1}(h, k, m) \geq c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h)$, i.e.,

$$\begin{aligned}
& (1 - \mathcal{D}(h, k, m)) \left(r(h, k, m) - c(h) - \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h) \right) \\
&\geq \lambda \sum_{h' \in S_H} v_n(h') (\mathcal{H}(h'|h) - \mathcal{Q}(h'|h)) \\
&\geq 0,
\end{aligned}$$

where the last inequality follows from [Lemma 2](#) and [Assumption 5](#). So,

$$r(h, k, m) - c(h) \geq \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h).$$

Since $\mathcal{D}(h, k, m) - \mathcal{D}(h+1, k, m) \leq 0$, by [Equation \(A5\)](#),

$$\begin{aligned}
& V_{n+1}(h, k, m) - V_{n+1}(h+1, k, m) \\
& \geq \mathbb{E}g((h, k, m), T) - \mathbb{E}g((h+1, k, m), T) \\
& + (\mathcal{D}(h, k, m) - \mathcal{D}(h+1, k, m))(r(h, k, m) - c(h)) \\
& = (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h) \\
& - (1 - \mathcal{D}(h+1, k, m))r(h+1, k, m) - \mathcal{D}(h+1, k, m)c(h+1) \\
& + (\mathcal{D}(h, k, m) - \mathcal{D}(h+1, k, m))(r(h, k, m) - c(h)) \\
& = (1 - \mathcal{D}(h+1, k, m))(r(h, k, m) - r(h+1, k, m)) \\
& + \mathcal{D}(h+1, k, m)(c(h) - c(h+1)) \\
& \geq 0,
\end{aligned}$$

where the last inequality follows from [Assumptions 1 and 2](#).

4. $V_{n+1}(h, k, m) = c(h) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h)$, $V_{n+1}(h+1, k, m) = (1 - \mathcal{D}(h+1, k, m)) \times r(h+1, k, m) + \mathcal{D}(h+1, k, m) (c(h+1) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h+1))$. Note that

$$\begin{aligned}
& (1 - \mathcal{D}(h+1, k, m))r(h+1, k, m) + \mathcal{D}(h+1, k, m)(c(h+1) \\
& + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h+1)) \\
& \geq c(h+1) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h+1).
\end{aligned}$$

Then,

$$\begin{aligned}
& (1 - \mathcal{D}(h+1, k, m)) \left(r(h+1, k, m) - c(h+1) - \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h+1) \right) \\
& \geq \lambda \sum_{h' \in S_H} v_n(h') (\mathcal{H}(h'|h+1) - \mathcal{Q}(h'|h+1)) \\
& \geq 0,
\end{aligned}$$

where the last inequality follows from [Lemma 2](#), [Assumption 5](#) and v_n being nonincreasing. So,

$$c(h+1) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h+1) \leq r(h+1, k, m). \quad (\text{A6})$$

Then,

$$\begin{aligned}
& V_{n+1}(h, k, m) - V_{n+1}(h+1, k, m) \\
& = \mathbb{E}g((h, k, m), T) + \mathcal{D}(h, k, m) \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h) \\
& - \mathbb{E}g((h+1, k, m), T) - \mathcal{D}(h+1, k, m) \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h+1) \\
& = \mathbb{E}g((h, k, m), T) - \mathbb{E}g((h+1, k, m), T) \\
& + \mathcal{D}(h, k, m) \lambda \sum_{h' \in S_H} v_n(h') (\mathcal{Q}(h'|h) - \mathcal{Q}(h'|h+1)) \\
& - (\mathcal{D}(h+1, k, m) - \mathcal{D}(h, k, m)) \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h+1) \\
& \geq \mathbb{E}g((h, k, m), T) - \mathbb{E}g((h+1, k, m), T) \\
& - (\mathcal{D}(h+1, k, m) - \mathcal{D}(h, k, m)) \lambda \sum_{h' \in S_H} v_n(h') \mathcal{Q}(h'|h+1) \\
& \geq \mathbb{E}g((h, k, m), T) - \mathbb{E}g((h+1, k, m), T) \\
& + (\mathcal{D}(h, k, m) - \mathcal{D}(h+1, k, m))(r(h+1, k, m) - c(h+1)) \\
& = (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h) \\
& - (1 - \mathcal{D}(h+1, k, m))r(h+1, k, m) - \mathcal{D}(h+1, k, m)c(h+1) \\
& + (\mathcal{D}(h, k, m) - \mathcal{D}(h+1, k, m))(r(h+1, k, m) - c(h+1)) \\
& = (1 - \mathcal{D}(h, k, m))(r(h, k, m) - r(h+1, k, m)) + \mathcal{D}(h, k, m)(c(h) - c(h+1)) \\
& \geq 0,
\end{aligned}$$

where the first inequality follows from [Lemma 2](#), [Assumption 5](#) and v_n being nonincreasing, and the second inequality follows from [Equation \(A6\)](#).

Therefore, $V_{n+1}(h, k, m)$ is nonincreasing in h . [Lemma 3](#) implies that v_{n+1} is nonincreasing in h . Finally, function V , as the limit of sequence $\{V_n\}$, is nonincreasing in both h and k , and [Lemma 3](#) implies that v is nonincreasing in h . ■

Proof of Theorems 3 and 4. To show the existence of the match-based control limit policy, it is equivalent to show that $T \in A^*(h, k, m+1)$ implies $T \in A^*(h, k, m)$ for any h, k, m , where $A^*(h, k, m)$ is the set of optimal actions at state (h, k, m) . Assume that $T \in A^*(h, k, m+1)$, then

$$\begin{aligned}
V(h, k, m+1) & = (1 - \mathcal{D}(h, k, m+1))r(h, k, m+1) \\
& + \mathcal{D}(h, k, m+1)(c(h) + \lambda \sum_{h' \in S_H} v(h') \mathcal{Q}(h'|h)) \\
& \geq c(h) + \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h),
\end{aligned}$$

i.e.,

$$\begin{aligned}
& (1 - \mathcal{D}(h, k, m+1))(r(h, k, m+1) - c(h)) \\
& \geq (1 - \mathcal{D}(h, k, m+1)) \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h) \\
& + \mathcal{D}(h, k, m+1) \lambda \sum_{h' \in S_H} v(h') (\mathcal{H}(h'|h) - \mathcal{Q}(h'|h)),
\end{aligned}$$

i.e.,

$$\begin{aligned}
& (1 - \mathcal{D}(h, k, m+1))(r(h, k, m+1) - c(h) - \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h)) \\
& \geq \mathcal{D}(h, k, m+1) \lambda \sum_{h' \in S_H} v(h') (\mathcal{H}(h'|h) - \mathcal{Q}(h'|h)) \\
& \geq 0,
\end{aligned}$$

where the last inequality follows from [Lemma 2](#), [Assumption 5](#) and v being nonincreasing. So, if $T \in A^*(h, k, m+1)$,

$$r(h, k, m+1) - c(h) - \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h) \geq 0. \quad (\text{A7})$$

Since $\mathcal{D}(h, k, m)$ is nondecreasing in m and $r(h, k, m)$ is nonincreasing in m , we have

$$\begin{aligned}
& (1 - \mathcal{D}(h, k, m))(r(h, k, m) - c(h) - \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h)) \\
& \geq \mathcal{D}(h, k, m) \lambda \sum_{h' \in S_H} v(h') (\mathcal{H}(h'|h) - \mathcal{Q}(h'|h)),
\end{aligned}$$

i.e.,

$$\begin{aligned}
V(h, k, m) & = (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)(c(h) \\
& + \lambda \sum_{h' \in S_H} v(h') \mathcal{Q}(h'|h)) \\
& \geq c(h) + \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h).
\end{aligned}$$

Therefore, $T \in A^*(h, k, m)$. The proof of [Theorem 4](#) is exactly the same (by showing $T \in A^*(h, k+1, m) \Rightarrow T \in A^*(h, k, m)$) and is omitted. ■

Proof of Corollary 1. If $m \geq M^*(h, k)$, it is optimal to choose W , and $V(h, k, m) = c(h) + \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h)$, as a function of m , is constant.

If $m < M^*(h, k)$, $T \in A^*(h, k, m)$ and

$$\begin{aligned}
V(h, k, m) & = (1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)(c(h) \\
& + \lambda \sum_{h' \in S_H} v(h') \mathcal{Q}(h'|h)) \\
& \geq c(h) + \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h).
\end{aligned}$$

In particular, $V(h, k, M^*(h, k) - 1) \geq V(h, k, M^*(h, k))$. From [Equation \(A7\)](#), we know that $r(h, k, m) - c(h) - \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h) \geq 0$ for $m < M^*(h, k)$.

Since $\sum_{h' \in S_H} v(h') \mathcal{H}(h'|h) \geq \sum_{h' \in S_H} v(h') \mathcal{Q}(h'|h)$ (follows from [Lemma 2](#), [Assumption 5](#) and v being nonincreasing), we have $r(h, k, m) - c(h) - \lambda \sum_{h' \in S_H} v(h') \mathcal{Q}(h'|h) \geq 0$ for $m < M^*(h, k)$. Rewrite

$$V(h, k, m) = (1 - \mathcal{D}(h, k, m))(r(h, k, m) - c(h)) - \lambda \sum_{h' \in S_H} v(h') \mathcal{Q}(h'|h) + \left(c(h) + \lambda \sum_{h' \in S_H} v(h') \mathcal{Q}(h'|h) \right).$$

Since $\mathcal{D}(h, k, m)$ is nondecreasing in m and $r(h, k, m)$ is nonincreasing in m , $V(h, k, m)$ is nonincreasing in m for $m < M^*(h, k)$. Thus, $V(h, k, m)$ is nonincreasing in m . ■

Proof of Corollary 2. When proving Theorem 3, we showed that for any $k, h, T \in a(h, k, m+1)$ implies that $T \in a(h, k, m)$, i.e., $k < K^*(h, m+1)$ implies that $k < K^*(h, m)$. Hence, $K^*(h, m+1) \leq K^*(h, m)$. Similarly, we can argue that $M^*(h, k+1) \leq M^*(h, k)$. The last two statements follow from monotonicity. ■

Lemma 4 is provided in Alagoz et al. (2007b).

Lemma 4. Suppose transition probability function $P(\cdot|\cdot) : S \times S \mapsto [0, 1]$ on state space $S = \{1, \dots, n\}$ is stochastically increasing. If function $f : \mathbb{R} \mapsto \mathbb{R}_+$ is nonincreasing, the following inequalities hold: for $i = 1, \dots, n, j = 1, \dots, n-1$,

1. $\sum_{i \leq j} (P(i|j) - P(i|j+1)) f(i) \geq \sum_{i \leq j} (P(i|j) - P(i|j+1)) f(j);$
2. $\sum_{i > j} (P(i|j) - P(i|j+1)) f(i) \geq \sum_{i > j} (P(i|j) - P(i|j+1)) f(j+1).$

Proof of Theorem 5. It is equivalent to show that $T \in A^*(h, k, m)$ implies that $T \in A^*(h+1, k, m)$ for any h, k , and m . Fix $h \leq H, k \leq K$. Prove by contradiction: suppose that $T \in A^*(h, k, m)$ but $A^*(h+1, k, m) = W$. Since $T \in A^*(h, k, m)$, we have

$$V(h, k, m) = \mathbb{E}g((h, k, m), T) + \mathcal{D}(h, k, m) \lambda \sum_{h' \in S_H} v(h') \mathcal{Q}(h'|h) \geq c(h) + \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h). \quad (\text{A8})$$

Since $A^*(h+1, m, k) = W$, the following strict inequality holds:

$$\mathbb{E}g((h+1, k, m), T) + \mathcal{D}(h+1, k, m) \lambda \sum_{h' \in S_H} v(h') \mathcal{Q}(h'|h+1) < c(h+1) + \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h+1). \quad (\text{A9})$$

Subtracting Equation (A9) from Equation (A8),

$$\begin{aligned} & \mathbb{E}g((h, k, m), T) + \mathcal{D}(h, k, m) \lambda \sum_{h' \in S_H} v(h') \mathcal{Q}(h'|h) \\ & - \mathbb{E}g((h+1, k, m), T) - \mathcal{D}(h+1, k, m) \lambda \sum_{h' \in S_H} v(h') \mathcal{Q}(h'|h+1) \\ & > c(h) - c(h+1) + \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h) - \lambda \sum_{h' \in S_H} v(h') \mathcal{H}(h'|h+1). \end{aligned}$$

Since $\mathcal{D}(h+1, k, m) \geq \mathcal{D}(h, k, m)$ (Assumption 3),

$$\begin{aligned} & \mathbb{E}g((h, k, m), T) - \mathbb{E}g((h+1, k, m), T) \\ & > c(h) - c(h+1) + \mathcal{D}(h, k, m) \lambda \sum_{h' \in S_H} v(h') (\mathcal{Q}(h'|h+1) - \mathcal{Q}(h'|h)) \\ & + \lambda \sum_{h' \in S_H} v(h') (\mathcal{H}(h'|h) - \mathcal{H}(h'|h+1)) \\ & \geq \mathcal{D}(h, k, m) \lambda \sum_{h' \in S_H} v(h') (\mathcal{Q}(h'|h+1) - \mathcal{Q}(h'|h)) \\ & + \lambda \sum_{h' \in S_H} v(h') (\mathcal{H}(h'|h) - \mathcal{H}(h'|h+1)) \\ & = \mathcal{D}(h, k, m) \lambda \sum_{h' \in S_H} v(h') (\mathcal{Q}(h'|h+1) - \mathcal{H}(h'|h+1) - (\mathcal{Q}(h'|h) - \mathcal{H}(h'|h))) \\ & + \lambda(1 - \mathcal{D}(h, k, m)) \sum_{h' \in S_H} v(h') (\mathcal{H}(h'|h) - \mathcal{H}(h'|h+1)) \\ & \geq \lambda(1 - \mathcal{D}(h, k, m)) \sum_{h' \in S_H} v(h') (\mathcal{H}(h'|h) - \mathcal{H}(h'|h+1)), \end{aligned} \quad (\text{A10})$$

where the second inequality follows from Assumption 2 that $c(h) \geq c(h+1)$; to prove the last inequality, it is

enough to notice that

$$\begin{aligned} & \sum_{h' \in S_H} v(h') (\mathcal{Q}(h'|h+1) - \mathcal{H}(h'|h+1) - (\mathcal{Q}(h'|h) - \mathcal{H}(h'|h))) \\ & = \sum_{h' \in S_H} v(h') (\mathcal{Q}(h'|h+1) + \mathcal{H}(h'|h) - (\mathcal{H}(h'|h+1) + \mathcal{Q}(h'|h))) \\ & \geq 0 \end{aligned}$$

because of Lemma 1, Assumption 9 and v being nonincreasing.

From Equation (A10), it follows that

$$\begin{aligned} & \mathbb{E}g((h, k, m), T) - \mathbb{E}g((h+1, k, m), T) \\ & > \lambda(1 - \mathcal{D}(h, k, m)) \sum_{h' \in S_H} v(h') (\mathcal{H}(h'|h) - \mathcal{H}(h'|h+1)) \\ & = \lambda(1 - \mathcal{D}(h, k, m)) \left(\sum_{h'=1}^h v(h') (\mathcal{H}(h'|h) - \mathcal{H}(h'|h+1)) \right. \\ & \quad \left. + \sum_{h'=h+1}^H v(h') (\mathcal{H}(h'|h) - \mathcal{H}(h'|h+1)) \right) \end{aligned} \quad (\text{A11})$$

(note that $v(H+1) = 0$). Since v is nonincreasing, by applying both parts of Lemma 4 to each of the sums in Equation (A11), respectively, we have

$$\begin{aligned} & \mathbb{E}g((h, k, m), T) - \mathbb{E}g((h+1, k, m), T) \\ & > \lambda(1 - \mathcal{D}(h, k, m)) \left(v(h) \sum_{h'=1}^h (\mathcal{H}(h'|h) - \mathcal{H}(h'|h+1)) \right. \\ & \quad \left. + v(h+1) \sum_{h'=h+1}^H (\mathcal{H}(h'|h) - \mathcal{H}(h'|h+1)) \right) = \lambda(1 - \mathcal{D}(h, k, m)) v(h) \\ & \quad \times \left(1 - \sum_{h'=h+1}^H \mathcal{H}(h'|h) - \mathcal{H}(H+1|h) \right) \\ & \quad - \left(1 - \sum_{h'=h+1}^H \mathcal{H}(h'|h+1) - \mathcal{H}(H+1|h+1) \right) \\ & \quad + \lambda(1 - \mathcal{D}(h, k, m)) v(h+1) \sum_{h'=h+1}^H (\mathcal{H}(h'|h) - \mathcal{H}(h'|h+1)) \\ & = \lambda(1 - \mathcal{D}(h, k, m)) (v(h) - v(h+1)) \sum_{h'=h+1}^H (\mathcal{H}(h'|h+1) - \mathcal{H}(h'|h)) \\ & \quad + \lambda(1 - \mathcal{D}(h, k, m)) v(h) (\mathcal{H}(H+1|h+1) - \mathcal{H}(H+1|h)). \end{aligned}$$

By Assumption 7, $\sum_{h'=h+1}^H (\mathcal{H}(h'|h+1) - \mathcal{H}(h'|h)) \geq 0$. Therefore,

$$\begin{aligned} & \mathbb{E}g((h, k, m), T) - \mathbb{E}g((h+1, k, m), T) \\ & > \lambda(1 - \mathcal{D}(h, k, m)) v(h) (\mathcal{H}(H+1|h+1) - \mathcal{H}(H+1|h)) \end{aligned} \quad (\text{A12})$$

From Assumption 8 and Equation (A12), we have

$$v(h) < \mathbb{E}g((h+1, k, m), T).$$

Since $\mathbb{E}g((h+1, k, m), T) \leq V(h+1, k, m)$, we obtain $v(h) < V(h+1, k, m)$. By Equation (A9),

$V(h+1, k, m) = c(h+1) + \lambda \sum_{h' \in S_H} v_n(h') \mathcal{H}(h'|h+1) = V(h+1, K+1, m)$. Then, by the monotonicity of $V(h, k, m)$ in both h and k ,

$$\begin{aligned} v(h) & < V(h+1, k, m) = V(h+1, K+1, m) \\ & \leq V(h, K+1, m) \leq v(h), \end{aligned}$$

where the last inequality follows from the fact that $V(h, K+1, m) \leq V(h, k, m), \forall k, m$, which is a contradiction. Therefore, $T \in A^*(h, k, m)$ implies that $T \in A^*(h+1, k, m)$. ■

Proof of Theorem 6. Prove by induction. Suppose that we solve Π_1 and Π_2 simultaneously using value iteration Equation (2) with $V_0^1(h, k, m) = V_0^2(h, k, m) = 0, \forall h, k$, and

m . Let V_j^i be the value function of $\Pi_i, i = 1, 2$ at the j^{th} iteration.

Initial step: at the first iteration,

$$\begin{aligned} V_1^1(h, k, m) &= V_1^2(h, k, m) \\ &= \begin{cases} \max((1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h), c(h)) & \text{if } k \leq K, \\ c(h) & \text{if } k = K. \end{cases} \end{aligned}$$

Thus, we have $U_1^1(h, k) = U_1^2(h, k), \forall h, k$. As shown in the proof of [Theorem 2](#), $V_1^1(h, k, m)$ and $V_1^2(h, k, m)$ are nonincreasing in h and k . Hence, $U_1^1(h, k)$ and $U_1^2(h, k)$ are nonincreasing in h and k as well.

Since $\mathcal{K}_2 \geq_{st} \mathcal{K}_1$, by [Lemma 1](#),

$$\begin{aligned} &\sum_{k \in S_K} (U_1^1(h, k)\mathcal{K}_1(k|h) - U_1^2(h, k)\mathcal{K}_2(k|h)) \\ &= \sum_{k \in S_K} U_1^1(h, k)(\mathcal{K}_1(k|h) - \mathcal{K}_2(k|h)) \geq 0, \forall h \in S_H, \end{aligned}$$

i.e., $v_1^1(h) \geq v_1^2(h), \forall h \in S_H$.

Induction step: now assume that $v_n^1(h) \geq v_n^2(h)$ and $V_n^1(h, k, m) \geq V_n^2(h, k, m)$ for all h, k , and m . We want to show $V_{n+1}^1(h, k, m) \geq V_{n+1}^2(h, k, m)$ for all h, k , and m . By inspecting [Equation \(2\)](#), we note that it suffices to show

$$\sum_{h' \in S_H} v_n^1(h')\mathcal{H}(h'|h) \geq \sum_{h' \in S_H} v_n^2(h')\mathcal{H}(h'|h), \quad (\text{A13})$$

$$\sum_{h' \in S_H} v_n^1(h')\mathcal{Q}(h'|h) \geq \sum_{h' \in S_H} v_n^2(h')\mathcal{Q}(h'|h). \quad (\text{A14})$$

We will show [Equation \(A13\)](#), and [Equation \(A14\)](#) can be proved in the same way. We write $v_n^i(h) = \sum_{k \in S_K} U_n^i(h, k) \times \mathcal{K}_i(k|h), i = 1, 2$. Since $V_n^1(h, k, m) \geq V_n^2(h, k, m)$ for all h, k , and m , $U_n^1(h, k) \geq U_n^2(h, k)$ for all h and k . Then,

$$\begin{aligned} &\sum_{h' \in S_H} v_n^1(h')\mathcal{H}(h'|h) - \sum_{h' \in S_H} v_n^2(h')\mathcal{H}(h'|h) \\ &= \sum_{h' \in S_H} \left(\sum_{k \in S_K} U_n^1(h', k)\mathcal{K}_1(k|h') \right) \mathcal{H}(h'|h) \\ &\quad - \sum_{h' \in S_H} \left(\sum_{k \in S_K} U_n^2(h', k)\mathcal{K}_2(k|h') \right) \mathcal{H}(h'|h) \\ &= \sum_{h' \in S_H} \mathcal{H}(h'|h) \sum_{k \in S_K} (U_n^1(h', k)\mathcal{K}_1(k|h') - U_n^2(h', k)\mathcal{K}_2(k|h')) \\ &\geq \sum_{h' \in S_H} \mathcal{H}(h'|h) \sum_{k \in S_K} (U_n^2(h', k)\mathcal{K}_1(k|h') - U_n^2(h', k)\mathcal{K}_2(k|h')). \end{aligned}$$

As shown in the proof of [Theorem 2](#), both $V_n^1(h, k, m)$ and $V_n^2(h, k, m)$ are nonincreasing in k for any n . Thus, both $U_n^1(h, k)$ and $U_n^2(h, k)$ are nonincreasing in k . Since $\mathcal{K}_2 \geq_{st} \mathcal{K}_1$, by [Lemma 1](#),

$$\sum_{k \in S_K} (U_n^2(h', k)\mathcal{K}_1(k|h') - U_n^2(h', k)\mathcal{K}_2(k|h')) \geq 0, \forall h' \in S_H.$$

Therefore,

$$\sum_{h' \in S_H} v_n^1(h')\mathcal{H}(h'|h) \geq \sum_{h' \in S_H} v_n^2(h')\mathcal{H}(h'|h).$$

It follows that $V_{n+1}^1(h, k, m) \geq V_{n+1}^2(h, k, m), \forall h, k$, and m . Then, $U_{n+1}^1(h, k) \geq U_{n+1}^2(h, k)$ for all h and k , and

$$\begin{aligned} &v_{n+1}^1(h) - v_{n+1}^2(h) \\ &= \sum_{k \in S_K} (U_{n+1}^1(h, k)\mathcal{K}_1(k|h) - U_{n+1}^2(h, k)\mathcal{K}_2(k|h)) \\ &\geq \sum_{k \in S_K} (U_{n+1}^2(h, k)\mathcal{K}_1(k|h) - U_{n+1}^2(h, k)\mathcal{K}_2(k|h)) \\ &\geq 0, \forall h \in S_H, \end{aligned}$$

where the second inequality follows from [Lemma 1](#) and

$\mathcal{K}_2 \geq_{st} \mathcal{K}_1$. Therefore, $v_n^1(h) \geq v_n^2(h)$ and $V_n^1(h, k, m) \geq V_n^2(h, k, m), \forall n$. Taking $n \rightarrow \infty$, we have $v^1(h) \geq v^2(h)$ and $V^1(h, k, m) \geq V^2(h, k, m), \forall h, k$, and m . ■

To prove [Theorem 7](#), we need the following lemma from Alagoz et al. (2007b).

Lemma 5. Suppose transition probability functions $P(\cdot|\cdot), Q(\cdot|\cdot)$ on state space $S = \{1, \dots, n\}$ satisfying $Q \geq_{st} P$. Then the following inequalities hold for any function non-increasing $f: \mathbb{R} \mapsto \mathbb{R}_+$: for $i, j = 1, \dots, n$,

1. $\sum_{j \leq i} (P(j|i) - Q(j|i))f(j) \geq \sum_{j \leq i} (P(j|i) - Q(j|i))f(i)$.
2. $\sum_{j > i} (P(j|i) - Q(j|i))f(j) \geq \sum_{j > i} (P(j|i) - Q(j|i))f(i+1)$.

Proof of Theorem 7. Prove by induction. Suppose that we solve Π_1 and Π_2 simultaneously using value iteration [Equation \(2\)](#) with $V_0^1(h, k, m) = V_0^2(h, k, m) = 0, \forall h, k$, and m . Let V_j^i be the value function of $\Pi_i, i = 1, 2$ at the j^{th} iteration.

Initial step: at the first iteration,

$$\begin{aligned} V_1^1(h, k, m) &= V_1^2(h, k, m) \\ &= \begin{cases} \max((1 - \mathcal{D}(h, k, m))r(h, k, m) + \mathcal{D}(h, k, m)c(h), c(h)) & \text{if } k \leq K, \\ c(h) & \text{if } k = K. \end{cases} \end{aligned}$$

Thus, $v_1^1(h) = v_1^2(h), \forall h$.

Induction step: assume that $V_n^1(h, k, m) \geq V_n^2(h, k, m), v_n^1(h) \geq v_n^2(h), \forall h, k$, and m . We want to show $V_{n+1}^1(h, k, m) \geq V_{n+1}^2(h, k, m), v_{n+1}^1(h) \geq v_{n+1}^2(h), \forall h, k$, and m . By inspecting [Equation \(2\)](#), we note that it suffices to show

$$\sum_{h' \in S_H} v_n^1(h')\mathcal{H}_1(h'|h) \geq \sum_{h' \in S_H} v_n^2(h')\mathcal{H}_2(h'|h), \quad (\text{A15})$$

$$\sum_{h' \in S_H} v_n^1(h')\mathcal{Q}_1(h'|h) \geq \sum_{h' \in S_H} v_n^2(h')\mathcal{Q}_2(h'|h). \quad (\text{A16})$$

We will show [Equation \(A15\)](#), and [Equation \(A16\)](#) can be proved in the same way. We have

$$\begin{aligned} &\sum_{h' \in S_H} v_n^1(h')\mathcal{H}_1(h'|h) - \sum_{h' \in S_H} v_n^2(h')\mathcal{H}_2(h'|h) \\ &\geq \sum_{h' \leq h} v_n^1(h')\mathcal{H}_1(h'|h) + \sum_{h' > h} v_n^2(h')\mathcal{H}_1(h'|h) \\ &\quad - \sum_{h' \leq h} v_n^2(h')\mathcal{H}_2(h'|h) - \sum_{h' > h} v_n^2(h')\mathcal{H}_2(h'|h) \\ &= \sum_{h' \leq h} v_n^2(h')(\mathcal{H}_1(h'|h) - \mathcal{H}_2(h'|h)) + \sum_{h' > h} v_n^2(h')(\mathcal{H}_1(h'|h) - \mathcal{H}_2(h'|h)) \\ &\geq v_n^2(h) \sum_{h' \leq h} (\mathcal{H}_1(h'|h) - \mathcal{H}_2(h'|h)) + v_n^2(h+1) \sum_{h' > h} (\mathcal{H}_1(h'|h) - \mathcal{H}_2(h'|h)) \\ &= (v_n^2(h) - v_n^2(h+1)) \sum_{h' \leq h} (\mathcal{H}_1(h'|h) - \mathcal{H}_2(h'|h)) \\ &\geq 0, \end{aligned} \quad (\text{A17})$$

where the first inequality follows from the induction assumption that $v_n^1(h') \geq v_n^2(h'), \forall h'$, the second inequality follows from [Lemma 5](#) and v_n^2 being nonincreasing (shown in the proof of [Theorem 2](#)), and the second equality follows from the fact that

$$\begin{aligned} &\sum_{h' \leq h} (\mathcal{H}_1(h'|h) - \mathcal{H}_2(h'|h)) + \sum_{h' > h} (\mathcal{H}_1(h'|h) - \mathcal{H}_2(h'|h)) \\ &= 0. \end{aligned}$$

The last equality holds because $v_n^2(h) \geq v_n^2(h+1)$ and $\sum_{h' \leq h} (\mathcal{H}_1(h'|h) - \mathcal{H}_2(h'|h)) \geq 0$, which follows from $\mathcal{H}_2 \geq_{st} \mathcal{H}_1$. Therefore, $V_{n+1}^1(h, k, m) \geq V_{n+1}^2(h, k, m), \forall h, k, m$.

Since Π_1 and Π_2 have the same \mathcal{K} and \mathcal{M} , $v_{n+1}^1(h) \geq v_{n+1}^2(h)$. Taking $n \rightarrow \infty$, we have $v^1(h) \geq v^2(h)$ and $V^1(h, k, m) \geq V^2(h, k, m), \forall h, k$, and m . ■

Appendix B. Selection of parameters in numerical experiments

B.1. Definition of states and transition law

The raw EPTS score is computed by the following formula (OPTN, 2022c):

$$\begin{aligned} \text{raw EPTS} = & (0.047 - 0.015 \times \mathbf{1}\{\text{diabetes}\}) \times (\text{age} - 25)^+ \\ & + (0.398 - 0.237 \times \mathbf{1}\{\text{diabetes}\}) \\ & \times \text{number of prior organ transplantations} \\ & + (0.315 - 0.099 \times \mathbf{1}\{\text{diabetes}\}) \\ & \times \log(\text{years on dialysis} + 1) \\ & + (0.130 - 0.348 \times \mathbf{1}\{\text{diabetes}\}) \times \mathbf{1}\{\text{no dialysis}\} \\ & + 1.262 \times \mathbf{1}\{\text{diabetes}\} \end{aligned}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function of an event and $x^+ := \max(x, 0)$. Patients with lower EPTS scores are expected to have longer time of graft function from high-longevity kidneys, compared to patients with higher EPTS scores. The raw EPTS is converted to an EPTS score (ranging from 0 to 100) using the EPTS mapping table. An online calculator can be found at the OPTN website (OPTN, 2022b).

To define the probability of death, i.e., $\mathcal{H}(H + 1|h)$, at each patient state, we find in the latest OPTN data report (OPTN, 2022a) that deaths per 100 patient years for kidney registrations during waiting is 6.91 for patients over 65, thus, the probability of death in a year is roughly 0.0691, and therefore, we use 0.035 as the probability of death for each epoch. We want to choose $\mathcal{H}(H + 1|h), \forall h$ to keep the arithmetic average of $\mathcal{H}(H + 1|h)$ over S_H close to 0.035. Since older patients are more likely to die, we take $\mathcal{H}(H + 1|h)$ to be increasing in h .

If the patient chooses to wait and is alive, their state transition law is deterministic as shown in Table B1 (transition probability function \mathcal{H} can be defined accordingly). If the patient experiences a transplantation failure, we define the state transition law according to the change in their EPTS score, as shown in Table B2 (transition probability function \mathcal{Q} can be defined accordingly). It is easy to check that Assumptions 4, 5 and 7, but not Assumption 9, are satisfied. However, Assumption 9 would hold if we directly define patient state by EPTS score, as Table B2

Table B1. Definition of patient state h and corresponding EPTS scores.

Patient state h	EPTS score
1	53
2	60
3	67
4	72
5	76
6	80
7	83
8	86
9	89
10	91
11	93
12	94
13	96
14	97
15	98
16	99 or greater
17	death

Table B2. Patient state transition law after a transplantation failure.

Pre-transplant EPTS score	Patient state h	Post-transplant EPTS score	Patient state h
53	1	79	6
60	2	85	8
67	3	89	9
72	4	92	10
76	5	95	12
80	6	97	13
83	7	98	14
86 or greater	8 or greater	99 or greater	16

Table B3. The distribution of kidney state.

KDPI range	Kidney state k	Probability (%)
0–20	1	4.91
21–34	2	3.23
35–85	3	12.06
86–100	4	3.47
Not available	5	76.53

Table B4. The distribution of mismatch level \mathcal{M} .

Mismatch level m	Probability (%)
1	4.92
2	1.04
3	1.92
4	14.37
5	28.06
6	32.54
7	14.14

indicates that there is a larger drop in EPTS score when patients with lower EPTS scores experience a transplant failure, compared with patients with higher EPTS scores.

We use the kidney donor profile index (KDPI) score to represent the **kidney state** k and assume that kidney state $\{k_n\}_{n=1}^\infty$ form an i.i.d. sequence of random variables, independent of both patient state and mismatch level. KDPI, a scalar that combines ten donor factors including clinical parameters and demographics, is used in the kidney allocation system to measure the quality of deceased donor kidneys (OPTN, 2022d). KDPI score ranges from 0 to 100. The average waiting time for a kidney offer is 2.13 years according to the latest OPTN report (OPTN, 2022a), so we assume that the patient's waiting time is a geometric random variable with mean 2.13 years. Therefore, the probability that a patient gets an offer is 0.2347 for each epoch (six months). We define four types of kidney state by their KDPI ranges. \mathcal{K} , the **distribution of the kidney state**, is obtained from OPTN (2022a) and shown in Table B3. Assumption 6 holds, because we assume that the distribution of kidney state is independent of patient state.

The **distribution of mismatch level** \mathcal{M} in deceased donor kidney transplantation, also obtained from OPTN (2022a), is given in Table B4.

We assume in Section 4 that the **probability of a transplantation failure** depends only on k and m , not on h , and we denote it by $\mathcal{D}(k, m)$. We use the six-month post-transplantation graft failure rate to represent $\mathcal{D}(k, m)$. For deceased donor kidney transplantation, the six-month post-transplantation graft survival rate is 97.1% for perfect match, and 94.7% for non-perfect match (OPTN, 2022a).

For deceased donor kidney transplantation, the six-month post-transplantation graft survival rate for different KDPI ranges is given in Table B6.

Table B5. Six-month post-transplantation graft survival rate for zero and nonzero mismatch levels (OPTN, 2022a).

Mismatch level m	Six-month graft survival probability (%)
$m = 1$	97.1
$m > 1$	94.7

Table B6. Six-month post-transplantation graft survival rate for different KDPI ranges (OPTN, 2022a).

KDPI range	Kidney state k	Graft survival probability (%)
0–20	1	97.1
21–34	2	95.1
35–85	3	94.1
86–100	4	91.6

Table B7. The probability (%) of a transplantation failure $\mathcal{D}(k, m)$.

$\mathcal{D}(k, m)$	$m = 1$	$m > 1$
$k = 1$	1.7	4.1
$k = 2$	3.7	6.1
$k = 3$	4.7	7.1
$k = 4$	7.3	9.5

Table B8. The 5-year post-transplantation patient survival rate (%).

EPTS \ KDPI	0–20	21–34	35–85	86–100
53–54	87.5	86.8	83.85	76.55
59–60	86	85.3	82.2	74.65
67–68	83.75	83.05	79.75	72
71–72	82.4	81.75	78.35	70.5
75–76	80.95	80.3	76.8	68.85
79–80	79.3	78.6	75.05	67.05
83–84	77.6	76.8	73.05	65.15
85–86	76.65	75.85	71.05	64.05
89–90	74.45	73.8	69.85	61.95
91–92	73.65	72.7	68.7	60.8
93–94	72.6	71.6	67.45	59.6
95–96	71.5	70.4	66.15	58.4
97	70.6	69.5	65.2	57.5
98	70	68.8	64.5	56.8
99 or greater	69.4	68.2	63.8	56.2

For each KDPI range, we assume that the six-month post-transplantation graft survival rate for $m = 1$ and $m > 1$ have the same ratio as Table B5, and the six-month graft survival rate for each KDPI range in Table B6 is equal to the arithmetic average of the survival rate for $m = 1$ and $m > 1$. Then, the probability of a transplantation failure $\mathcal{D}(k, m)$ is given in Table B7. We verify that Assumption 3 holds for \mathcal{D} specified in Table B7.

B.2. Rewards

We set the **intermediate reward** to be 0.5 years, i.e., $c(h) = 0.5, \forall h$. We use the expected post-transplantation survival time to represent the **post-transplantation reward** $r(h, k, m)$. We assume that the patient post-transplantation survival time is a Poisson random variable (with unit to be a year). If we know the five-year post-transplantation patient survival rate, we can approximate the expected post-transplantation survival time with the mean of the corresponding Poisson random variable. For each EPTS-KDPI pair, the five-year post-transplantation patient survival rate can be found in Bae et al., (2019). For each EPTS-KDPI range in Table B8, the five-year post-transplantation survival rate is approximated by the arithmetic average of survival rate at endpoints of

Table B9. Relative risk for different HLA-mismatch levels.

Mismatch level m	Relative risk
1	0.9
2	1
3	1.1
4	1.2
5	1.3
6	1.4
7	1.6

Table B10. Expected post-transplantation patient survival time for $m = 1$.

EPTS \ KDPI	0–20	21–34	35–85	86–100
53–54	12	11	10	8.5
59–60	11	11	9.5	8.2
67–68	9.9	9.7	9	7.9
71–72	9.6	9.4	8.8	7.7
75–76	9.3	9.1	8.5	7.6
79–80	8.9	8.8	8.3	7.4
83–84	8.7	8.5	8	7.2
85–86	8.5	8.4	7.8	7.1
89–90	8.3	8.1	7.7	6.9
91–92	8.1	8	7.5	6.8
93–94	8	7.9	7.4	6.7
95–96	7.8	7.7	7.3	6.6
97	7.7	7.6	7.2	6.6
98	7.7	7.6	7.1	6.5
99 or greater	7.6	7.5	7.1	6.5

Table B11. Expected post-transplantation patient survival time for $m = 7$.

EPTS \ KDPI	0–20	21–34	35–85	86–100
53–54	6	5.9	5.8	5.5
59–60	5.9	5.9	5.8	5.5
67–68	5.8	5.8	5.7	5.4
71–72	5.8	5.7	5.6	5.3
75–76	5.7	5.7	5.6	5.3
79–80	5.6	5.6	5.5	5.2
83–84	5.6	5.6	5.4	5.1
85–86	5.5	5.5	5.3	5.1
89–90	5.5	5.4	5.3	5
91–92	5.4	5.4	5.3	5
93–94	5.4	5.4	5.2	4.9
95–96	5.4	5.3	5.2	4.9
97	5.3	5.3	5.1	4.8
98	5.3	5.2	5.1	4.8
99 or greater	5.3	5.2	5.1	4.8

the EPTS range with KDPI score to be the median of the KDPI range.

To compute the post-transplantation reward, we also need to take into consideration the effect of mismatch level. We use the relative risk (Opelz and Döhler, 2007) to measure relative contribution of HLA to the five-year patient survival rate.

We set mismatch level $m = 5$ to be the reference and use the relative risk measure shown in Table B9. Given EPTS score, KDPI score, and mismatch level, the five-year patient survival rate is obtained by the corresponding probability in Table B8 divided by the relative risk in Table B9.

Table B10 and Table B11 show the **expected post-transplantation patient survival time** for $m = 1$ and $m = 7$, respectively. When the EPTS or KDPI score is low, the expected post-transplantation survival time is much longer for $m = 1$, compared with $m = 7$. When the EPTS or KDPI score is high, the effect of mismatch level is less obvious. Assumptions 1 and 2 hold with the reward functions we used.