Stochastic Modeling of Multistate Disease Dynamics under Random Environments

M. Manoharan
Department of Statistics
University of Calicut
Kerala-673 635, India.
manumavila@gmail.com

T.D. Xavier
Department of Statistics
St.Thomas College
Thrissur, Kerala - 680 001, India.
xavierstc@yahoo.co.in

***Abstract*—This paper presents the application of Semi-Markov decision Process (SMDP) for a multi-state disease under random environments to determine the optimal treatment strategy. The subject/patient lives in varying random environments, imparting significant effects on performance/health status. While the environment evolves according to a Semi-Markov Process, in each environment state, the subject goes through several states of disease according to a Semi-Markov Process. In an environment 'k' when the patient state is 'i', one of the following two actions are available: continue the present treatment strategy (C) with a given cost rate $h^k(i)$ or initiate a rejuvenating treatment strategy (R) with a cost rate $c^k(i)$. In this complex model the optimal strategy is found out minimizing the expected discounted total cost. A special case of Markov environment is discussed indicating the feasibility of the computation of optimal policy. A numerical illustration is also provided to support the viability of the analysis and results. The model provides a useful and flexible representation of acute and chronic events and can be used to explore the economic impact of changes in therapy.**

***Keywords*-Stochastic model; Semi-Markov Process; Semi-Markov Decision Process (SMDP); Multistate Disease Dyanamics; Random environment; Optimal Treatment strategy.**

I. Introduction

In some medical treatments, decision must be made sequential and in an uncertain environment. A physician determining a course of treatment must consider patient's health as well as the best treatment decision in the future. Often decisions are to be taken in a dynamic environment. Physiological as well as physical changes in patients, may sometime contribute to the changes of the environment. Uncertain environment arises mainly due to patients respond differently even to same treatment for a disease.

Physicians always need to make subjective judgement about the treatment strategies. However a mathematical decision model that provide insight into the nature of optimal decision can aid the treatment. Markov decision processes (MDPs) are appropriate technique useful in class of problems involving complex, stochastic and dynamic decisions for which it can find optimal solutions. The goal of a MDP is to provide a optimal policy which is a decision strategy to optimize a particular criterion such as maximizing total discounted reward. It not only provides consequence of a policy, but also guarantees that no better policy exists.

MDPs are a general framework for modeling dynamic systems under uncertainty. It binds previous, current and future treatment decision through the proper definition of patient's states defined as variables that contain the relevant information for making future decisions. The treatment model evolves in the following manner. The condition or state of the patient is observed (or partially observed), an action is taken, a cost is incurred (or a reward is received) and the patient gets into a new state according to a known probability distribution. The state variable is defined so that given the current state of patient, the future transitions and rewards are independent of past. This is the standard assumption of a Markov process.

The broad classes of MDPs are finite horizon MDPs and infinite horizon MDPs. The number N of decision epochs is finite in the former and it goes to infinite in the latter. For a finite horizon model optimal policy for both the average reward per state and the total reward criterion are equivalent. Infinite horizon models requires a large amount of data hence it is assumed that the data are time homogeneous. As a result the states of infinite horizon MDP must be carefully defined to ensure that the state transition of patients are stationary. When the data are time dependent the time homogeneity assumption can be satisfied by properly augmenting the state definition with the time at which the transition occurs.

In most of the medical investigations, the state of patients decided in the light of results from a series of medical tests are subjected to test errors. A modified MDPs called Partially Observed MDPS (POMDP) have been developed to deal the data with imperfect information ([1], [2]). In these models it is assumed that uncertainty exist, in patient's transition and the state he truly occupies. Therefore the objective is to find an optimal policy based on the observation of the patient and the previous decision rule applied. As reported in [3] a Markov decision approach has been employed for multi-category patient scheduling decisions in computed tomography (CT) and investigated associated tradeoffs from economic and operational perceptives. Risk-sensitive optimality criteria for MDPs have been considered by various authors over the years. Often they lead to non-standard MDPs involving time-inconsistency. In the last decade risk measures have been popular and the

978-1-4673-9941-8/16 $31.00 © 2016 IEEE
DOI 10.1109/SMRLO.2016.69

simple variance has been replaced by more complicated risk measures like Value-at-Risk (VaR) or Avearge-Value-at-Risk (AVaR). [4] has given a short account of this research problem and investigated the problem of minimizing the AVaR of the discounted cost over a finite and infinite horozon which is generated by a MDP. As reported recently in [5] the MDP has been applied in the agricultural sector to find optimal policy for managing an orange farm, which have been solved by policy iteration and linear programming.

In MDP models the treatment decision are taken at each of a sequence of unit time intervals or fixed epochs and the sojourn time in states has no effect on rewards or incurring costs for patient. However in health-care and other application, decision are taken over continuous time intervals such as varying treatment can be administered. The sojourn time in states may depend on the duration of his/her current health status. The MDP models might not be suitable to model such disease progression instead Semi-Markov Decision Process (SMDP) models are more appropriate. SMDP models allow the patients' state transition to occur in continuous time and facilitate to assume any probability distribution for sojourn time in a state.

Markov and Semi-Markov process are appropriate model for the multi-state diseases in which the data arises as transition-times and states. Historically it have been difficult to adopt realistic models for biomedical applications since the likelihood turns out to be real complicated. With the progress on numerical methods for estimation, problems of tractability can be overcome. [6] introduced a k-state Markov model for continuous time processes. Similar models have been applied to AIDS ([7]), heart transplantation ([8]), diabetes ([9], [10]), infectious diseases and cancer screening and so on. These models have been further extended to include fixed time and time-varying covariate information (see for eg. [11], [12]). Methods for estimation of transition rates are generally numerically based and have usual maximum likelihood sampling schemes such as Metroplois-Hasting method ([13], [14]) or have used population-based approaches similar to weighted least squares. These approaches have all concentrated on continuous-time Markov processes usually due to unequally spaced observation times.

In this paper, we consider a complex survival model that lives in a randomly changing environment which affect model parameters. The term '*environment*' is used in the generic sense so that it represents any set of conditions that affect the stochastic structure of the model investigated. The concept of environment process, in one form or another, has been used in the literature for various purposes. The use of environmental process to modulate the deterministic and stochastic parameters of Operation Research models can be seen in reliability, inventory and queueing applications. The problem of optimal replacement of a semi-Markov system under semi-Markov environment is studied by [15] (see also [16]). [17], [18] discuss other applications in inventory and queueing. A comprehensive discussion on Markov modulated queueing system can be found in [19].

Although the literature cited above illustrate the use of random environment in reliability, inventory and queueing model, the concept is of paramount interest in survival analysis. It is generally assumed that a patient stays in a given fixed environment. The probability law of his ageing and death process there remain intact throughout his useful life. The life duration and corresponding hazard rate is taken to be the one obtained through statistical life testing procedures that are believed to be under ideal conditions. There has been growing interest in recent years in lifetime models under random environment.

This is necessitated by the fact that the subject/patient often lives in varying environments during which they are subjected to varying environment conditions with significant effects on performance/health status. During a treatment period whole environment of the patient may change due to occurrence of other contagious diseases, hypertension, high blood pressure, cardiac problems, severe climatic/seasonal changes or adopting entirely new treatment strategy on medical team's advice. When environment changes, the state of patient also drastically changes. The deterioration and failure process therefore depends on the environment. This makes it crucial to identify an optimal treatment strategy especially for a range of multi-state disease process.

II. The Model

The general form of the model that represents the foregoing situation may be stated as follows:

1. The patient is in a semi-Markov environment $\{(J_n, L_n), n \geq 0\}$ on a set K of countable environment states, where J_n is the state of the environment immediately after its n^{th} transition epoch L_n, and $0 = L_0 < L_1 < L_2 < \cdots$. Clearly $L_{n+1} - L_n$ is the time duration of the patient in the state J_n. Let the state's kernel be
$$G_{kk'}(t) = Pr(L_{n+1} - L_n \leq t, J_{n+1} = k'/J_n = k)$$
and let $\psi_{kk'} = G_{kk'}(\infty)$ and $G_k(t) = \sum_{k' \in K} G_{kk'}(t)$.
2. During an environment state k, the patient goes through several states of disease according to a semi-Markov process with a kernel $\{P^k_{ij}(t), i, j \in S\}$ and a set $S = \{0, 1, 2, ...\}$ of countable states, where the state 0 represents disease-free state, and states 1, 2, . . . represent the different adverse disease states of the patient and the bigger the value, the more serious is the condition.
Let $P^k_{ij} = P^k_{ij}(\infty), T^k_{ij}(t) = P^k_{ij}(t)/P^k_{ij}$ and $T^k_i(t) = \sum_{j \in S} P^k_{ij}(t)$
3. Suppose that the patient is in environment k, then one of the following two actions can be chosen if his state transfers to i :
(a) Continue the present treatment strategy (denoted by

C) with a cost rate $h^k(i)$;
(b) Initiating a rejuvenating treatment strategy (denoted by R) like chemotherapy, radiation, surgery, organ transplant and/or admission in ICU etc., with a cost rate $c^k(i)$,
and the time of action R is assumed to be a random variable with probability distribution function $F^k(t)$, and the state after rejuvenation will be 0, the disease-free state.

4. When the environment state changes from k to k' if action C or R is chosen then the patient's state will change immediately according to a probability q^k_{ij} and an instantaneous cost $R^k(i, C)$ incurs; while if action R is chosen then within no time it completed and an instantaneous cost $R^k(i, R)$ incurs.
5. The objective is to minimize the expected discounted total costs with discount factor $\alpha > 0$.

The above treatment strategy can be modeled by a semi-Markov decision process (SMDP) in a semi- Markov environment as follows.

During the environment state k, i.e., $J_n = k$ for some $n \geq 0$, it can be modeled by the following SMDP:

$$SMDP_k := \{S, A, P^k(j|i, a), T^k(.|i, a, j), r^k(i, a, j, u)\} \tag{1}$$

where S is the state space and A ={C, R}is the action set. The transition probability P^k, the distribution function T^k of the transition time, and the one step cost function r^k are given, respectively by

$$\begin{aligned} P^k(j|i, C) &= P^k_{ij} \\ P^k(j|i, R) &= \delta_{i0} \\ T^k(t|i, C, j) &= T^k_{ij}(t) \\ T^k(t|i, R, 0) &= F^k(t) \end{aligned}$$

$$\begin{aligned} r^k(i, C, j, u) &= h^k(i) \int_0^u e^{-\alpha t} dt = h^k(i)\alpha^{-1}(1 - e^{-\alpha t}) \\ r^k(i, R, j, u) &= C^k(i) \int_0^u e^{-\alpha t} dt = C^k(i)\alpha^{-1}(1 - e^{-\alpha t}) \end{aligned} \tag{2}$$

$$\begin{aligned} where\ \ \delta_{i0} &= 1,\ \text{if } j = 0 \\ &= 0,\ \text{otherwise.} \end{aligned}$$

For SMDP in a semi-Markov environment, when the environment state changes from k to k' i.e., at L_{n+1} for some $n \geq 0$ with $J_n = k$ and $J_{n+1} = k'$, the patient's state changes immediately to j with a probability $q(j|i, a, k, k')$ if the patient's state is i at $L_{n+1} - 0$ and the last action taken before L_{n+1} is 'a', and in the same time, an instantaneous cost $R^k(i, a)$ occurs, where

$$q(j|i, C, k, k') = q^k_{ij} \text{ and } q(j|i, R, k, k') = \delta_{j0}.$$

To simplify notations, for $k \in K$ and s, t ≥ 0, we let

$$\begin{aligned} h^k(s, t) &= \sum_{k'} P(L_{n+1} - L_n > t, J_{n+1} = k'/J_n = k) \\ &\int_0^t e^{-\alpha u} du \\ &+ \int_{s^+}^{s+t} \sum_k P(L_{n+1} - L_n \leq u, J_{n+1} = k'/J_n = k) \\ &\int_0^{u-s} e^{-\alpha l} dl \\ &= \alpha^{-1}(1 - e^{-\alpha t})[1 - G_k(t + s)] + \\ &\alpha^{-1} \int_{s^+}^{s+t} (1 - e^{-\alpha(u-s)} dG_k(u), \\ g^{kk'}(s, t) &= \int_{s^+}^{s+t} e^{-\alpha(u-s)} dG_{k'}(u), \\ g^k(s, t) &= \sum_{k' \in K} g^{k.k'}(s, t) = \int_{s^+}^{s+t} e^{-\alpha(u-s)} dG_k(u). \end{aligned} \tag{3}$$

Let $x = (k, s, i) \in \Omega = \{(k, s, i) : k \geq 0, s \geq 0, i \in S\}$ be the mathematical state which means that the environment is in state k just since time s ago and the patient's state just transfers to i . Then, by denoting the expected discounted cost occurring when the state x is reached and action 'a' is taken, as $r(x, a)$ we have

$$\begin{aligned} r(x, C) &= h^k(i) \int_0^\infty h^k(s, t) dT^k_i(t) \\ &\quad + R^k(i, C) \int_0^\infty g^k(s, t) dT^k_i(t) \\ r(x, R) &= c^k(i) \int_0^\infty h^k(s, t) dF^k(t) \\ &\quad + R^k(i, R) \int_0^\infty g^k(s, t) dF^k(t) \end{aligned} \tag{4}$$

and $\beta(x, a, k')$ is corresponding to a discount factor depending on the state x,when the action is 'a' and the next environment state k',

$$\begin{aligned} \beta(x, C, k') &= \int_0^\infty g^{kk'}(s, t)\ dT^k_i(t) \\ \beta(x, R, k') &= \int_0^\infty g^{kk'}(s, t)\ dF^k(t). \end{aligned}$$

Now, it follows that $V^*(x)$, the minimal expected discounted total cost starting from the initial state x, is the minimal nonnegative solution of the following optimality equation

$$V^*(x) = min\{V^*(x, C), V^*(x, R)\} \tag{5}$$

where, $x=(k,s,i)\in\Omega$ and

$$V^*(x,C)=r(x,C)+\sum_{k'\in K}\beta(x,C,k')\sum_{j\in S}q_{ij}^k\ V^*(k',0,j)$$
$$+\sum_{j\in S}P_{ij}^k\int_0^\infty e^{-\alpha t}V^*(k,s+t,j)\ dT_{ij}^k(t) \qquad (6)$$
$$V^*(x,R)=r(x,R)+\sum_{k'\in K}\beta(x,R,k')\ V^*(k',0,0)$$
$$+\int_0^\infty e^{-\alpha t}V^*(k,s+t,0)\ dF^k(t)$$

are respectively the discounted total cost if action C or R is used in the first horizon with the mathematical state x and then the optimal policy is used in the remaining horizons.

III. Optimal Control Limit Policies

From the standard results in discrete time Markov decision processes DTMDP, Eqn.(5) can be considered as an optimality equation for an adequate DTMDP with state space Ω. Thus we can consider its n-horizon problem with the optimality equation

$$V_n^*(x)=min\{V_n^*(x,C),V_n^*(x,R)\}, \text{for } x=(k,s,i)\in\Omega \qquad (7)$$

where $V_n^*(x)$ is the optimal value from state x for n horizons problem, while

$$V_n^*(x,C)=r(x,C)+\sum_{k\in K}\beta(x,C,k')\sum_{j\in S}q_{ij}^k\ V_{n-1}^*(k',0,j)$$
$$+\sum_{j\in S}P_{ij}^k\int_0^\infty e^{-\alpha t}V_{n-1}^*(k,s+t,j)\ dT_{ij}^k(t) \qquad (8)$$
$$V_n^*(x,R)=r(x,R)+\sum_{k\in K}\beta(x,R,k')\ V_{n-1}^*(k',0,0)$$
$$+\int_0^\infty e^{-\alpha t}V_{n-1}^*(k,s+t,0)\ dF^k(t)$$

are the values from state x in n horizons if action C or R is used respectively in the first horizon and then an optimal policy in the remaining horizons. The initial conditions are

$$V_0^*(x,C)=V_0^*(x,R)=0$$

. Let

$$v_n(x)=V_n^*(x,C)-V_n^*(x,R),$$
$$v(x)=V^*(x,C)-V^*(x,R),\ \ x=(k,s,i)\in\Omega$$

then by the standard theory in DTMDP

$$\lim_{n\to\infty}V_n^*(x,a)=V^*(x,a),\ \ a=C,R \qquad (9)$$
$$\lim_{n\to\infty}v_n(x)=v(x)$$

while the optimal policies can be depicted as

$$f_n^*(x)=C\Longleftrightarrow v_n(x)<0, f^*(x)=C\Longleftrightarrow v(x)<0.$$

Hence $(f_N^*,f_{N-1}^*,...,f_0^*)$ is optimal for N-horizon problem; and f^* is optimal for the infinite horizon discounted criterion. A concept of stochastic order between two distribution functions is needed. For two distribution functions F and G, F is said to be smaller stochastically than G, denoted by $F\preceq G$, if $F(t)\geq G(t)$ for each t.
We have the following familiar result on stochastic order.
For two distribution functions F and G, $F\preceq G$ if and only if $\int_{-\infty}^{\infty}f(t)dF(t)\leq\int_{-\infty}^{\infty}f(t)dG(t)$ for each nondecreasing function f .
To obtain some properties of the optimal policies, we make the following assumptions.
Assumptions A: For each $k\in K$,

- A.1 $\sum_{j=m}^{\infty}q_{ij}^k$ nondecreasing in i for each $m\geq 0$;
- A.2 $h^k(i),c^k(i),R^k(i,C)$ and $R^k(i,R)$ are all nonnegative and nondecreasing in i ;
- A.3 both $h^k(i)-c^k(i)$ and $R^k(i,C)-R^k(i,R)$ are nondecreasing in i ;
- A.4 $F^k\preceq T_0^k\preceq T_1^k\preceq T_2^k\ldots$, i.e., T_i^k is stochastically nondecreasing in i ;
- A.5 $\int_0^\infty e^{-at}\sum_{j\in S}V(t,j)p_{ij}^k dT_{ij}^k(t)$ is nondecreasing in i if V(t,j) is nonnegative and nondecreasing in j for each $t\ \geq\ 0$.

Note that Assumption (A.1) that $\sum_{j=m}^{\infty}q_{ij}^k$ is nondecreasing in i for $m>0$ means that more serious is the condition of patient, faster the critical state reached by the environment change. Assumption (A.3) that $h^k(i)-c^k(i)$ is nondecreasing in i indicates that the treatment cost increases faster than the rejuvenating cost as the condition of patient gets more serious and similarly for $R^k(i,C)-R^k(i,R)$. In fact, Assumptions (A.1), (A.2) and (A.3) are those in the literature for the discrete time model, while Assumption (A.4) is given for the continuous time case here. Assumption (A.4) means that the sojourn time in a state for patient is nondecreasing as seriousness of his condition increases, and the rejuvenating time is smaller than the sojourn time in any state. Assumption (A.5) follows that if $T_{ij}^k(t)$ is absolutely continuous with probability density function $t_{ij}^k(t)$, and $\sum_{j=m}^{\infty}p_{ij}^k t_{ij}^k(t)$ is nondecreasing in i for each $t\geq 0, m\geq 0$, which is similar as (A.1). It is easy to see that the latter two conditions are involved in the state definition.

By the earlier stated result on stochastic order and Assumption A, it is easy to see that the $g_{kk'}(t)$ is nondecreasing in i which implies that $\beta(x,C,k')$ is nondecreasing in i.

We have the following well known result on transition probabilities.
Let $[r_{ij}]$ be a transition probability matrix, then the following two are equivalent:
(1) For each $m\geq 0$, $\sum_{j=m}^{\infty}r_{ij}$ is nondecreasing in i ;
(2) For each nonnegative and nondecreasing function h(j), $\sum_{j=0}^{\infty}r_{ij}h(j)$ is nondecreasing in i.

Then by using the induction method it can be shown from the above result and Assumption A that all of $V_n^*(x,C)$,

$V_n^*(x,R)$ and $V^*(x)$ are nondecreasing in i.

Now for each $k \in K$ and $s \geq 0$,

$r(x,C) - r(x,R)$

$$= h^k(i)\int_0^\infty h^k(s,t)\,[dT_i^k(t) - dF^k(t)]$$
$$+ [h^k(i) - c^k(i)]\int_0^\infty h^k(s,t)\,dF^k(t)$$
$$+ R^k(i,C)\int_0^\infty g^k(s,t)\,[dT_i^k(t) - dF^k(t)]$$
$$+ [R^k(i,C) - R^k(i,R)]\int_0^\infty g^k(s,t)dF^k(t)$$

is also nondecreasing in i due to Assumption A, result on stochastic order and the fact that both $h^k(s,t)$ and $g^k(s,t)$ are nondecreasing in t for each $k \in K, s \geq 0$. It should be noted that the latter two terms in $V_n^*(x,R)$ of Eqn.(8) are independent of i. So $v_n(x)$ is nondecreasing in i and thus we have the following result.

Theorem 3.1: Under Assumptions A, both $V_n^*(k,s,i)$ and $v_n(k,s,i)$ are nondecreasing in i for each $n \geq 0, k \in K$, $s \geq 0$, so

$$\begin{aligned} V_n^*(k,s,i) &= V_n^*(k,s,i,c), 0 \leq i \leq i_n^*(k,s) \\ &= V_n^*(k,s,i,R), i \geq i_n^*(k,s) \qquad (10) \\ where\ \ i_n^*(k,s) &:= min(i|v_n(k,s,i) \geq 0). \end{aligned}$$

Similarly, both $V^*(k,s,i)$ and $v(k,s,i)$ are also nondecreasing in i and

$$\begin{aligned} V^*(k,s,i) &= V^*(k,s,i,c), 0 \leq i \leq i^*(k,s) \\ &= V^*(k,s,i,R), i \geq i^*(k,s) \qquad (11) \\ where\ \ i^*(k,s) &:= min(i|v(k,s,i) \geq 0). \end{aligned}$$

The above Theorem states that there exists a state limit $i^*(k,s)$ just since time s ago for each $k \in K$ and $s \geq 0$ such that if the patient enters a state i while the environment is in state k, then the optimal action is to start treatment R if and only if the deteriorative degree of the patient is over the limit $i^*(k,s)$, i.e., $i \geq i^*(k,s)$. Such a policy is called a control limit policy. So the Theorem shows that there exists optimal control limit policies for both finite and infinite-horizon problems.

In the next section, we will discuss a special case of Markov environment and get more illuminating results.

IV. Markov Environment

In this section, we consider that the environment is Markov as follows:

$$G_{kk'}(t) = \psi_{kk'}G_k(t), G_k(t) = 1 - e^{-\lambda_k t}, t \geq 0, k,\ k' \in K. \qquad (12)$$

In this case, it will be shown that the variable s in state x=(k,s,i) can be deleted.

Let

$$\begin{aligned} t_F^k &= \int_0^\infty [1 - e^{-(\lambda_k+\alpha)t}]dF^k(t) \\ t_{ij}^k &= \int_0^\infty [1 - e^{-(\lambda_k+\alpha)t}]dT_{ij}^k(t)] \\ t_t^k &= \sum_{j\in S} P_{ij}^k t_{ij}^k,\ \alpha_F^k = 1 - t_F^k,\ \alpha_{ij}^k = 1 - t_{ij}^k,\ \alpha_i^k = 1 - t_i^k \end{aligned} \qquad (13)$$

where $F^k(t)$ and $T_{ij}^k(t)$ are defined in Section II. Then it can be calculated, due to Eqn.(4), that

$$\begin{aligned} r(x,C) &= r'(k,i,C)e^{-(\lambda_k s)} \\ &= \frac{t_i^k}{\lambda_k + \alpha}[h^k(i) + \lambda_k R^k(i,C)]e^{-(\lambda_k s)} \\ r(x,R) &= r'(k,i,R)e^{-(\lambda_k s)} \\ &= \frac{t_F^k}{\lambda_k + \alpha}[c^k(i) + \lambda_k R^k(i,R)]e^{-(\lambda_k s)} \qquad (14) \\ \beta(x,C,k') &= \frac{\lambda_k t_i^k}{\lambda_k + \alpha}\psi_{kk'}e^{-\lambda_k s} \\ \beta(x,R,k') &= \frac{\lambda_k t_F^k}{\lambda_k + \alpha}\psi_{kk'}e^{-\lambda_k s}. \end{aligned}$$

Based on Eqn.(14), it can be shown that $e^{\lambda_k s}V^*(k,s,i)$ and therefore $e^{\lambda_k s}V^*(k,s,i,C)$, $e^{\lambda_k s}V^*(k,s,i,R)$ are independent of s and thus

$$\begin{aligned} e^{\lambda_k s}V^*(k,s,i) &= V^*(k,0,i) \\ e^{\lambda_k s}V^*(k,s,i,C) &= V^*(k,0,i,C) \\ e^{\lambda_k s}V^*(k,s,i,R) &= V^*(k,0,i,R) \end{aligned}$$

We denote by

$$V^*(k,i) := V^*(k,0,i),\ V^*(k,i,C) := V^*(k,0,i,C),$$
$$V^*(k,i,R) := V^*(k,0,i,R),\ and$$
$$v(k,i) = V^*(k,i,C) - V^*(k,i,R).$$

Then $V^*(k,i)$ is the minimal nonnegative solution of the following optimality equation

$$V^*(k,i) = min\{V^*(k,i,C),\ \ V^*(k,i,R)\} \qquad (15)$$

with corresponding

$$\begin{aligned} V^*(k,i,C) &= r'(k,i,C) + \frac{\lambda_k t_i^k}{\lambda_k + \alpha}\sum_{k'\in K}\psi_{kk'}\sum_{j\in S} q_{ij}^k V^*(k',j) \\ &+ \sum_{j\in S} p_{ij}^k \alpha_{ij}^k V^*(k,j) \\ V^*(k,i,R) &= r'(k,i,R) + \frac{\lambda_k t_F^k}{\lambda_k + \alpha}\sum_{k'\in K}\psi_{kk'}V^*(k',0) \\ &+ \alpha_F^k V^*(k,0) \end{aligned} \qquad (16)$$

Now, the problem is simplified by deleting the time variable s, and we can solve for $V^*(k,i)$ only. From the standard results in DTMDP, Eqn.(15) can also be considered as the optimality equation of an adequately defined DTMDP with state space $S' = \{(k,i) : k \in K, i \in S\}$ and action set A ={C, R}.

In the case of Markov environment, Assumptions (A.4) and (A.5) can be replaced, respectively, by the following weaker ones:

A.4′ $t_F^k \le t_0^k \le t_1^k \le t_2^k \le \ldots$ for each $k \in K$;

A.5′ $\sum_{j=m}^{\infty} p_{ij}^k \alpha_{ij}^k$ is nondecreasing in i for each $k \in K$ and $m \ge 0$.

The following corollary can be proved on similar lines as in Theorem 3.1 based on the above discussions.

Corollary 4.1: For the Markov environment case, suppose that Assumptions (A.1), (A.2), (A.3), (A.4′), (A.5′) hold, then $v(k,i) = V^*(k,i,C) - V^*(k,i,R)$ is nondecreasing in i and

$$\begin{aligned} V^*(k,i) &= V^*(k,i,C), \quad i < i^*(k) \\ &= V^*(k,i,R), \quad i \geqslant i^*(k) \end{aligned}$$

where $i^*(k,i) = min\{i|v(k,i) \ge 0\}$.

The corollary says that the state limit is also independent of the time variable s, that is, $i^*(k,s) = i^*(k)$.

Remark 4.1: (1) If $t_F^k \le t_0^k$ is not true, then it can be shown similarly that Corollary 4.1 holds in $i \ge J_k := min\{i|t_i^k \ge t_F^k\}$ and $i^*(k)$ should be redefined by $i^*(k) := min\{i \ge J_k|\, v(k,i) \ge 0\}$. In this case, the optimal policy is to do with (C) if $J_k \le i < i^*(k)$ and with (R) if $i \ge i^*(k)$, while it is not known what optimal action is when $0 \le i < J_k$. We call such a policy an extended control limit policy.

(2) Due to the ex-pressions of $r(x,a)$ of Eqn. 14, one can know that both the optimal value and the optimal policies depend on $T_{ij}^k(t)$ only through t_{ij}^k. This is to say that the model with a Markov environment is robust with respect to the distribution function $T_{ij}^k(t)$ of the time of state transition for the patient.
Moreover, if

$$P_{ij}^k(t) = P_{ij}^k T_i^k(t), \quad \forall\, i,j,k$$

then, we can assume that the patient is Markov, i.e.,

$$T_i^k(t) = 1 - e^{-\mu_i^k t}$$

where μ_i^k and t_i^k are determined by each other by

$$t_i^k = \frac{\lambda_k + \alpha}{\lambda_k + \alpha + \mu_i^k} \ ,$$

$$\mu_i^k = (\lambda_k + \alpha)\frac{1 - t_i^k}{t_i^k}.$$

In Assumption (A.4′), t_i^k is nondecreasing in i, so $\sum_{j=m}^{\infty} P_{ij}^k \alpha_{ij}^k = \sum_{j=m}^{\infty} P_{ij}^k (1 - t_{ij}^k)$ may not be nondecreasing. The following lemma gives a sufficient condition for it.

Lemma 4.1: Suppose that $T_{ij}^k(t) = T_i^k(t)$ for all $i,j \in$ S, k $\in$ K and m ≥ 0, then $\sum_{j=m}^{\infty} p_{ij}^k \alpha_{ij}^k$ is nondecreasing in i if and only if

$$\frac{t_{i+1}^k - t_i^k}{1 - t_i^k} \le \frac{\sum_{j=m}^{\infty} P_{i+1,j}^k - \sum_{j=m}^{\infty} P_{i,j}^k}{\sum_{j=m}^{\infty} P_{ij}^k} \qquad (17)$$

(Equation(17) means that the increasing speed of $\sum_{j=m}^{\infty} P_{ij}^k$ in i for each m ≥ 0 is larger than or equal to the decreasing speed of $(1 - t_i^k)$.)

Proof: It follows from the given condition that

$$t_{i,j}^k = t_i^k,\ \alpha_{i,j}^k = \alpha_i^k = 1 - t_i^k,\ \sum_{j=m}^{\infty} P_{ij}^k \alpha_{ij}^k = (1 - t_i^k) \sum_{j=m}^{\infty} P_{ij}^k$$

It is obvious that for two nonnegative functions h(i) and g(i), if h(i) is non-increasing while g(i) is nondecreasing, then h(i),g(i) is nondecreasing if and only if

$$\frac{h(i)}{h(i+1)} \le \frac{g(i+1)}{g(i} \ or \ \frac{h(i) - h(i+1)}{h(i+1)} \le \frac{g(i+1) - g(i)}{g(i}$$

which immediately implies the Lemma 4.1. □

The optimal policies f_n^* and f^* are characterized by $i_n^*(k)$ and $i^*(k)$, respectively. We have the following result about the upper bound of these numbers, which is useful for the state reduction problem discussed below.

Lemma 4.2: Under the conditions given in Corollary 4.1, if $t_0^k = t_F^k$, then $i_n^*(k) \le i_0^*(k) := min\{i|\triangle r(k,i) \ge 0\}$ and$i^*(k) \le i_0^*(k)$ where $\triangle r(k,i) = r'(k,i,C) - r'(k,i,R)$.

Proof: If $t_0^k = t_F^k$, then $\alpha_0^k = \alpha_F^k$. So it follows from earlier results that

$$\begin{aligned} & t_i^k \sum_{j\,\in\, S} q_{ij}^k V_n(k',j) - t_F^k V_n(k',0) \\ \ge\ & t_F^k \sum_{j\,\in\, S} q_{ij}^k V_n(k',0) - t_F^k V_n(k',0) \ = \ 0 \text{ and} \\ & \sum_{j\,\in\, S} p_{ij}^k \alpha_{ij}^k V_n(k,j) - \alpha_F^k V_n(k,0) \\ \ge\ & \sum_{j\,\in\, S} p_{0j}^k \alpha_{0j}^k V_n(k,j) - \alpha_F^k V_n(k,0) \\ =\ & \alpha_0^k V_n(k,0) - \alpha_F^k V_n(k,0) = 0 \end{aligned}$$

So, we get $v_n(k,i) \ge \triangle r(k,i)$, which implies the assertion of the lemma. □

Assumption (A.3) is about the cost rate, we now replace it by a new one about the expected total cost in a state.

(A.3′): for each k $\in$ K, both $h^k(i)t_i^k - c^k(i)t_F^k$ and $R^k(i,C)t_i^k - R^k(i,R)t_F^k$ are nondecreasing in i .

Here, $h^k(i)t_i^k$ and $c^k(i)t_F^k$ are respectively the expected treatment and rejuvenating treatment costs in state i when the environment state is k. So the nondecreasing of $h^k(i)t_i^k - c^k(i)t_F^k$ means that the expected treatment cost increases faster than the expected rejuvenating treatment cost as the patient's state increases. The nondecreasing of $R^k(i,C)t_i^k - R^k(i,R)t_F^k$ has a similar meaning.

Theorem 4.1: Under Assumptions (A.1), (A.2), (A.3′), (A.4′) and (A.5′), for each k $\in$ K and $n \geq 1, v_n(k,i) := V_n^*(k,i,C) - V_n^*(k,i,R)$ is nondecreasing in i, so $v_n(k,i) < 0$ if, and only if, $i < i_n^*(k) := min\{i | v_n(k,i) \geq 0\}$. Moreover, $v(k,i) := V^*(k,i,C) - V^*(k,i,R)$ is also nondecreasing in i, and $v(k,i) < 0$ if, and only if, $i < i^*(k) := min\{i | v(k,i) \geq 0\}$. That is, there exist optimal control limit policies.

Proof: First we note that under the given conditions,

$$(\lambda_k + \alpha)\triangle r(k,i) = [h^k(i)t_i^k - c^k(i)t_F^k] + \lambda_k[R^k(i,C)t_i^k - R^k(i,R)t_F^k]$$

is nondecreasing in i. Then the theorem can be proved exactly as that of Theorem 3.1. □

Remark 4.2: The above theorem shows the existence of optimal control limit policies whose state limit $i^*(k)$ depends only on the environment state k. Thus, the Markov environment case is more simpler than the semi-Markov environment case.

Now we reduce the number of states of the patient under the Markov environment (see Eqn.(12)). First, we suppose that

$$i_n^*(k) \leq j(k),\ n \geq 0,\ k \in K \tag{18}$$

for some j(k), where $i_n^*(k)$ is defined in Theorem 4.1. By Theorem 4.1, we have

$$V^*(k,i) = V^*(k,i,R) = r'(k,i,R) + V_0(k), i \geq j(k), k \in K \tag{19}$$

where

$$V_0(k) = \frac{\lambda_k t_F^k}{\lambda_k + \alpha} \sum_{k' \in K} \psi_{kk'} V^*(k',0) + \alpha_F^k V^*(k,0).$$

Thus one may get by Eqn.(15) for i $\geq$ 0 as follows:

$$\begin{aligned} V^*(k,i,C) &= r'(k,i,C) \\ &= + \frac{\lambda_k t_i^k}{\lambda_k + \alpha} \sum_{k' \in K} \psi_{kk'} \{ \sum_{j=0}^{j(k')-1} q_{ij}^k V^*(k',j) \\ &+ \sum_{j=j(k')}^{\infty} q_{ij}^k [r'(k,i,R) + V_0(k')]\} \\ &+ \sum_{j=0}^{j(k)-1} p_{ij}^k \alpha_{ij}^k V^*(k,j) \\ &+ \sum_{j=j(k)}^{\infty} p_{ij}^k \alpha_{ij}^k [r'(k,i,R) + V_0(k)] \end{aligned}$$

After manipulating the terms we get,

$$\begin{aligned} V^*(k,i,C) &= r'(k,i,C) \frac{\lambda_k t_i^k}{\lambda_k + \alpha} \sum_{k' \in K} \psi_{kk'} \\ & \sum_{j=j(k')}^{\infty} q_{ij}^k [r'(k',j,R) - r'(k',j(k'),R)] \\ & + \sum_{j=j(k')}^{\infty} p_{ij}^k \alpha_{ij}^k [r'(k,j,R) - r'(k,j(k),R)] \\ & + \frac{\lambda_k t_i^k}{\lambda_k + \alpha} \sum_{k' \in K} \psi_{kk'} \\ & [\sum_{j=0}^{j(k')-1} q_{ij}^k V^*(k',j) + \sum_{j=j(k')}^{\infty} q_{ij}^k V^*(k',j(k'))] \\ & + \sum_{j=0}^{j(k)-1} p_{ij}^k \alpha_{ij}^k V^*(k,j) + \sum_{j=j(k)}^{\infty} p_{ij}^k \alpha_{ij}^k V^*(k,j(k)) \end{aligned} \tag{20}$$

We define that

$$\begin{aligned} \widetilde{q}_{ij}^{kk'} &= q_{ij}^k, & j < j(k') \\ &= \sum_{j=j(k')}^{\infty} q_{ij}^k & j = j(k') \\ \widetilde{p}_{ij}^k &= p_{ij}^k, & j < j(k') \\ &= \sum_{j=j(k')}^{\infty} p_{ij}^k & j = j(k') \end{aligned}$$

$$\begin{aligned} \widetilde{T}_{ij}^k(t) &= T_{ij}^k(t), & j < j(k') \\ &= \sum_{j=j(k)}^{\infty} p_{ij}^k T_{ij}^k(t) / \widetilde{p}_{i,j(k)}^k & j = j(k') \end{aligned}$$

Thus

$$\sum_{j=j(k)}^{\infty} p_{ij}^k \alpha_{ij}^k = \widetilde{p}_{i,j(k)}^k \widetilde{\alpha}_{i,j(k)}^k$$

where $\widetilde{\alpha}_{i,j(k)}^k$ is defined as $\alpha_{i,j(k)}^k$ with $T_{ij}^k(t)$ being replaced by $\widetilde{T}_{ij}^k(t)$. Let

$$\widetilde{h}^k(i) = h^k(i) + \lambda_k \sum_{k'\in K} \psi_{kk'} \sum_{j=j(K')}^{\infty} q_{ij}^k \frac{t_F^{k'}}{\lambda_{k'}+\alpha}[c^{k'}(j) - c^{k'}(j(k'))] + (t_i^k)^{-1} t_F^k \sum_{j=j(K')}^{\infty} p_{ij}^k \alpha_{ij}^k [c^{k'}(j) - c^{k'}(j(k'))]$$

$$\widetilde{R}^k(i,C) = R^k(i,C) + \sum_{k'\in K} \psi_{kk'} \sum_{j=j(k')}^{\infty} q_{ij}^k \frac{t_F^{k'}}{\lambda_{k'}+\alpha} \lambda_{k'} x [R^{k'}(j,R) - R^{k'}(j(k'),R)] + (t_i^k)^{-1} t_F^k \sum_{j=j(K)}^{\infty} p_{ij}^k \alpha_{ij}^k [R^k(j,R) - R^k(j(k),R)]$$

$$\widetilde{r}(k,i,C) = \frac{t_i^k}{\lambda_k+\alpha}[\widetilde{h}^k(i) + \lambda_k R^k(i,C)]. \tag{21}$$

It is easy to see that $\widetilde{r}(k,i,C)$ is still nondecreasing in i for each k under Assumption A. Then for $i \geq 0$,

$$V^*(k,i,C) = \widetilde{r}(k,i,C) + \frac{\lambda_k t_i^k}{\lambda_{k'}\alpha} \sum_{k'\in K} \psi_{kk'} \sum_{j=0}^{j(k')} \widetilde{q}_{ij}^{kk'} V^*(k',j) + \sum_{j=0}^{j(k)} \widetilde{p}_{ij}^k \widetilde{\alpha}_{ij}^k V^*(k,j)$$

Now, we construct a new rejuvenating model (NRM), which is similar as the original rejuvenating model (ORM) excepts that

1 the state set of the patient in environment k is $S_k = \{0, 1, ..., j(k)\}$ for k ∈ K;
2 the parameters p_{ij}^k, $T_{ij}^k(t)$, q_{ij}^k, $h^k(i)$ and $R^k(i,C)$ are replaced by $\widetilde{p}_{ij}^k$,
$\widetilde{T}_{ij}^k(t)$, $\widetilde{q}_{ij}^k$, $\widetilde{h}^k(i)$ and $\widetilde{R}^k(i,C)$, respectively, which are defined the above;
3 the patient must be given rejuvenating treatment in state $j(k)$ during environment state k (due to Eqn.(18).

From the above discussions, we know that the NRM and the ORM are equivalent under the meanings that the optimal objective values are identical and their optimality equations are equivalent for both the finite- and infinite-horizon problems. So their optimal policies are identical. The difference between them is that the number of patient's states is finite for NRM. Certainly, the problem with finite states is simpler than that with infinite states, e.g., the computation for the case of finite states is feasible while that for the case of infinite states should be approximated.

When $j(k) \leq j^*$ for some j^*, it can take the state set as $S_k = \{0, 1, ..., j^*\}$, which is irrespective of k.

In what follows, we consider two further special cases. The first is that the state of patient itself is Markov, i.e.,

$$T_{ij}^k(t) = 1 - e^{-\mu_{k.i}t} \qquad F^k(t) = 1 - e^{-\mu_F(t} \tag{22}$$

then

$$\alpha_{ij}^k = \frac{\mu_{k.i}}{\lambda_k + \mu_{k.i} + \alpha} \quad \& \quad t_{ij}^k = \frac{\lambda_k + \alpha}{\lambda_k + \mu_{k.i} + \alpha}.$$

The second further special case is that the environment is a Poisson process with rate λ, i.e., the Markov environment (see Eqn.(12) with

$$\psi_{k,k'} = 1 \ , \ G_k(t) = 1 - e^{-\lambda t}, \ \ t \geq 0, \ \ k \in K. \tag{23}$$

Moreover, it is assumed that each adverse factor increases the degree of the seriousness in the condition of the patient with a probability distribution $\{q_j, j \geq 0\}$ as follows:

$$q_{ij}^k = 0 \ \ for\ j < 0 \quad and \ \ q_{ij}^k = q_{j-1} \ \ for\ j \geq i. \tag{24}$$

Furthermore, all $p_{ij}^k, T_{ij}^k, h^k(i), c^k(i), R^k(i,C)$ and $R^k(i,R)$ are independent of k and will be denoted by p_{ij} , T_{ij} and so on, by only deleting k in the original notations. Then $t_{ij}^k, t_i^k, \alpha_{ij}^k, \alpha_i^k, t_F^k, \alpha_F^k$ are also independent of k and will be denoted by t_{ij} , t_i and so on.
Under these conditions, it can be shown that $V^*(k,i)$ and therefore $V^*(k,i,C)$ and $V^*(k,i,R)$ are independent of k. So $i^*(k) = i^*$ is also independent of k.

V. Numerical Example

Consider a numerical example where the environment is a Markov process having two states with parameters as follows

$$\psi_{kk'} = \begin{pmatrix} 0.64 & 0.36 \\ 0.57 & 0.43 \end{pmatrix}, \ \ \lambda_1 = 0.076, \ \ \lambda_2 = 0.093$$

and the state transition probabilities for the system are

$$(P_{ij}^1) = \begin{pmatrix} 0.68 & 0.26 & 0.06 & 0.0 & 0.0 \\ 0.04 & 0.68 & 0.22 & 0.00 & 0.0 \\ 0.0 & 0.07 & 0.65 & 0.28 & 0.00 \\ 0.0 & 0.0 & 0.03 & 0.65 & 0.32 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix},$$

$$(P_{ij}^2) = \begin{pmatrix} 0.83 & 0.15 & 0.02 & 0.0 & 0.0 \\ 0.03 & 0.72 & 0.18 & 0.07 & 0.0 \\ 0.0 & 0.04 & 0.67 & 0.18 & 0.11 \\ 0.0 & 0.0 & 0.05 & 0.61 & 0.34 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix},$$

while two probability systems caused by the environment changes are

$$(q^1_{ij}) = \begin{pmatrix} 0.58 & 0.28 & 0.14 & 0.00 & 0.0 \\ 0.06 & 0.48 & 0.28 & 0.18 & 0.0 \\ 0.0 & 0.04 & 0.41 & 0.27 & 0.18 \\ 0.0 & 0.0 & 0.0 & 0.42 & 0.58 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix},$$

$$(q^2_{ij}) = \begin{pmatrix} 0.64 & 0.26 & 0.10 & 0.0 & 0.0 \\ 0.06 & 0.55 & 0.27 & 0.12 & 0.0 \\ 0.0 & 0.06 & 0.47 & 0.29 & 0.18 \\ 0.0 & 0.0 & 0.0 & 0.48 & 0.52 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}$$

The cost rate functions are as follows:

$$h^1(i) = 18{+}2i, c^1(i) = 46{+}i, R^1(i, C) = 55{+}i, R^1(i, R) = 0$$

$$h^2(i) = 15{+}2i, c^2(i) = 41{+}i, R^2(i, C) = 50{+}i, R^2(i, R) = 0$$

Now, it is assumed that the continuous discount factor is α = 0.45, and

$$(t^1_F, t^1_1, t^1_2, t^1_3, t^1_4) = (0.46, 0.73, 0.74, 0.76, 0.78, 0.80),$$
$$(t^2_F, t^2_1, t^2_2, t^2_3, t^2_4) = (0.57, 0.78, 0.79, 0.80, 0.82, 0.86).$$

Thus for i= 0,1,2,3,4

$$r'(1, i, C) = \frac{t^1_i}{\lambda_1 + \alpha}(21.496 + 2.076i),$$
$$r'(2, i, C) = \frac{t^1_i}{\lambda_1 + \alpha}(18.731 + 2.093i),$$
$$r'(1, i, R) = \frac{t^1_F}{\lambda_1 + \alpha}(46 + i),$$
$$r'(2, i, R) = \frac{t^1_F}{\lambda_1 + \alpha}(41 + i).$$

Now we compute the finite-horizon optimal values $V_n(k, i)$ iteratively by

$$V_{n+1}(k, i, C) = r'(k, i, C) + \frac{\lambda_k t^k_i}{\lambda_{k'} + \alpha} \sum_{k' \in K} \psi_{kk'} \sum_{j \in S} q^k_{ij} V_n(k', j\ ) + \sum_{j \in S} p^k_{ij}(1 - t^k_j)V_n(k, j), \quad (25)$$

$$V_{n+1}(k, i, R) = r'(k, i, R) + \frac{\lambda_k t^F_i}{\lambda_{k'} + \alpha} \sum_{k' \in K} \psi_{kk'} V_n(k', 0) + (1 - t^k_F)V_n(k, 0), \quad (26)$$

$$V_{n+1}(k, i) = min\{V_{n+1}(k, i, C) - V_{n+1}(k, i, R)\}$$
$$for\ n\ \geq\ 0\ with\ V_0(k, i, C) = V_0(k, i, R) = 0,\ \forall\, k, i.$$

The numerical results are shown in tables I and II when n=26,
$|V_{n+1}(k, i, a) - V_n(k, i, a)|\ \leq\ 0.01$ for all (k,i,a),
so we take the optimal value $V^*(k, i) = V_{27}(k, i)$.

Table I
ENVIRONMENT 1 - VALUES OF $V_n(k, i)$ & $v(k, i)$

	$V_n(1, i)$				
n	$i = 0$	1	2	3	4
1	129.69	144.16	161.09	178.72	197.02
2	225.48	247.43	270.47	291.30	295.11
3	297.28	324.65	346.80	384.85	402.44
4	355.19	384.76	407.90	459.72	477.68
5	398.77	429.02	451.97	510.15	515.60
6	430.30	459.77	480.78	538.91	542.71
7	452.28	480.17	499.12	557.72	561.52
8	467.18	493.55	511.09	570.46	574.26
9	477.14	502.36	519.05	578.96	582.76
10	483.76	508.19	524.34	584.61	588.42
11	488.15	512.05	527.86	588.37	592.17
12	491.07	514.62	530.21	590.86	594.66
13	493.01	516.32	531.76	592.52	596.32
14	494.30	517.45	532.79	593.61	597.42
15	495.15	518.20	533.48	594.34	598.15
16	495.72	518.70	533.94	594.83	598.63
17	496.10	519.03	534.24	595.15	598.95
18	496.35	519.25	534.44	595.36	599.17
19	496.51	519.40	534.57	595.51	599.31
20	496.62	519.50	534.66	595.60	599.40
21	496.70	519.56	534.72	595.66	599.46
22	496.74	519.61	534.76	595.70	599.51
23	496.78	519.63	534.79	595.73	599.53
24	496.80	519.65	534.80	595.75	599.55
25	496.81	519.67	534.81	595.76	599.56
26	496.82	519.67	534.82	595.77	599.57
$v(k, i)$	-87.54	-68.49	-57.15	2.28	13.99

Table II
ENVIRONMENT 2 - VALUES OF $V_n(k, i)$ & $v(k, i)$

	$V_n(2, i)$				
n	$i = 0$	1	2	3	4
1	105.87	119.20	132.83	148.57	168.85
2	195.97	219.98	246.56	273.52	284.67
3	289.54	324.08	358.84	389.73	412.60
4	350.36	389.05	428.46	460.69	464.82
5	391.23	431.43	470.73	494.62	498.75
6	418.94	458.48	496.05	518.92	523.05
7	437.53	475.89	512.38	535.75	539.89
8	449.91	487.25	523.23	547.14	551.27
9	458.13	494.74	530.46	554.73	558.86
10	463.59	499.70	535.29	559.78	563.91
11	467.21	502.99	538.50	563.14	567.27
12	469.61	505.17	540.63	565.36	569.49
13	471.21	506.62	542.05	566.84	570.97
14	472.27	507.59	542.99	567.82	571.95
15	472.97	508.23	543.61	568.48	572.61
16	473.44	508.65	544.03	568.91	573.04
17	473.75	508.93	544.30	569.20	573.33
18	473.95	509.12	544.49	569.39	573.52
19	474.09	509.25	544.61	569.51	573.64
20	474.18	509.33	544.69	569.60	573.73
21	474.24	509.38	544.74	569.65	573.78
22	474.28	509.42	544.78	569.69	573.82
23	474.31	509.44	544.80	569.72	573.85
24	474.33	509.46	544.82	569.73	573.86
25	474.34	509.47	544.83	569.74	573.87
26	474.34	509.48	544.84	569.75	573.88
$v(k, i)$	-83.01	-52.01	-20.78	5.58	16.40

Now $v(k, i)$ is shown in the last lines of tables I and II and thus the optimal limits for both the environments 1 and

2 is 3. That is,

$$v^*(1) = 3, \quad v^*(2) = 3.$$

The optimal policy in this example, is to initiate a rejuvenating treatment strategy if and only if the state of the patient reaches or exceeds 3 in both the environment 1 and 2.

VI. Conclusion

In this paper, we studied an optimal treatment strategy of a patient in a semi-Markov environment. We considered the performance/health status of the patient in a semi-Markov setup and so be the influence of the environment to the patient. For both the finite- and infinite-horizon discounted criterions, it was shown that there exist optimal control limit policies. A special case for a Markov environment was discussed. When the control limits are bounded for each environment state, the countable states of the patient was simplified equivalently to a finite one. Finally, the results are illustrated by a numerical example, supporting the viability and validity of the analysis.

References

[1] W. S. Lovejoy, "A survey of algorithm methods for partially observed Markov decision problems," *Annals of Operation Research*, **28**, pp. 47-66, 1991.

[2] C. C. White and W. T. Scherer, "Solution procedures for partially observed Markov decision processes," *Operation Research* , **37**, pp. 791-797, 1989.

[3] Y. Gocgun, B. W. Bresnahan, A. Ghate and M. L.Gunn, "A Markov decision approach to multi-category patient scheduling in a diagnostic facility," *Artificial Intelligence in Medicine,* **53** , pp. 673-681, 2011.

[4] N. Bäuerle and J.Ott, "Markov decision processes with Average-Value-at-Risk criteria," *Math Meth Oper Res,* **74** , pp. 361-379, 2011.

[5] J. M. Izar-Landeta, C. B.Ynzunza-Cortes and H. M.Azua, "Markovian decision process to find optimal policies in the management of an orange farm," *Revista Investigacion Operacional,* **35** , No.1, pp. 68-77, 2014.

[6] R. Kay, "A Markov model for analyzing cancer markers and disease states in survival studies," *Biometrics,* **42** , pp. 855-865, 1986.

[7] I. M. Longini Jr, W. S. Clark and R. H. Byers, "Statistical analysis of the stages of HIV infection using a Markov model," *Stat, Med.,* **8**, pp. 831-843, 1989.

[8] L. D. Sharples, "Use of Gibbs sampler to estimate transition rates between grades of coronary disease following cardiac transplantation," *Stat Med,* **12** , pp. 1155-1169, 1993.

[9] P. K. Andersen, "Multi-state model in survival analysis: A study of nephropathy and mortality in diabetes," *Stat. Med.,* **7** , pp. 661-670, 1988.

[10] G. Marshall and R. H.Jones, "Multi-state models and diabetic retinopathy," *Stat. Med.,* **14**, pp. 1975-1983, 1995.

[11] P. K. Andersen , L. S. Hansen and N. Keiding, "Assessing the influence of reversible disease indicators on survival," *Stat Med*, **10**, pp. 1061-1067, 1991.

[12] K. Gauvreau, V. Degruttola and M. Pagano, "The effect of covariates on the induction time of AIDS using improved imputation of exact seroconversion times," *Stat. Med.* , **13** , pp. 2021-2030, 1994.

[13] T. C. Prevost, G. Launoy, S. W. Duffy and H. H. Chen, "Estimating sensitivity and sojourn time in screening colorectal cancer: A comparison of statistical approaches, *Am. J. Epidemiol.*," **148** , pp. 609-619, 1998.

[14] A. Richard, S. Richardson and J. Maccario, "A three-state Markov model of plasmodium falciparum parasitemia," *Math. Biosci.*, **117** , pp. 283-300, 1993.

[15] Q. Hu and W. Yue, "Optimal replacement of a system according to a semi-Markov decision process in a semi-Markov environment," *Optimization Methods and Software*, **18**, pp. 181-196, 2003.

[16] Q. Hu, "Discounted semi-Markov decision process in a semi-Markov environment", *Optimization*, **39**, pp. 367-382, 1997.

[17] S. Ozekici, "Complex Systems in Random Environments," in Reliability and Maintenance of Complex Systems, Edited by S. Ozekici, Berlin, *Springer-Verlag,* NATO ASI Services, Vol F154, 137-157, 1996.

[18] S. Ozekici and M. Parlar, "Inventory Models with Unreliable Suppliers in a Random Environment," *Annals of Operation Research*, **91**, pp. 123-136, 1999.

[19] N. V. Prabu and Y. Zhu, "Markov Modulated Queueing System," *Queueing Systems*, **5**, pp. 215-246, 1989.