A reinforcement learning-based optimal control approach for managing an elective surgery backlog after pandemic disruption

Huyang Xu[1] · Yuanchen Fang[2] · Chun-An Chou[3] · Nasser Fard[3] · Li Luo[2]

Abstract

Contagious disease pandemics, such as COVID-19, can cause hospitals around the world to delay nonemergent elective surgeries, which results in a large surgery backlog. To develop an operational solution for providing patients timely surgical care with limited health care resources, this study proposes a stochastic control process-based method that helps hospitals make operational recovery plans to clear their surgery backlog and restore surgical activity safely. The elective surgery backlog recovery process is modeled by a general discrete-time queueing network system, which is formulated by a Markov decision process. A scheduling optimization algorithm based on the piecewise decaying ϵ-greedy reinforcement learning algorithm is proposed to make dynamic daily surgery scheduling plans considering newly arrived patients, waiting time and clinical urgency. The proposed method is tested through a set of simulated dataset, and implemented on an elective surgery backlog that built up in one large general hospital in China after the outbreak of COVID-19. The results show that, compared with the current policy, the proposed method can effectively and rapidly clear the surgery backlog caused by a pandemic while ensuring that all patients receive timely surgical care. These results encourage the wider adoption of the proposed method to manage surgery scheduling during all phases of a public health crisis.

Keywords Pandemic disruption · Elective surgery backlog · Stochastic scheduling optimization · Queueing network system · Markov decision process · Reinforcement learning · Operations research · Operations management

Highlights

- A method for clearing and managing the backlog of elective surgeries following a pandemic is proposed.
- The elective surgery backlog recovery process is modeled by a queueing network system.
- A reinforcement learning-based backlog scheduling optimization algorithm is proposed.
- The proposed method can rapidly clear the elective surgery backlog while ensuring timely care for all patients.

✉ Yuanchen Fang
  y.fang@scu.edu.cn

1 College of Management Science, Chengdu University of Technology, Chengdu, Sichuan, China

2 Department of Industrial Engineering and Management, Business School, Sichuan University, Chengdu, Sichuan, China

3 Department of Mechanical & Industrial Engineering, Northeastern University, Boston, MA, USA

1 Introduction

The coronavirus disease 2019 (COVID-19) pandemic has caused hospitals around the world to temporarily delay nonemergent elective surgeries to reduce the infectious risk to both patients and providers while conserving hospitals' capacities and available resources to take care of those patients who are the sickest due to COVID-19 [1]. This delay of surgeries has resulted in a backlog of uncompleted procedures that had been previously scheduled, as well as a dynamic backlog of surgeries that continue to be delayed as the health system experiences diminished capacity [2]. Studies have estimated that more than 28 million surgeries were either canceled or postponed during the first peak 12 weeks of the COVID-19 pandemic [3]. Taking the United States as an example, even under the most optimistic scenario, the United States may face a cumulative backlog of more than a million total joint and spine surgery cases and 1.1 million to 1.6 million cataract procedures by 2022, and the country may need up to 16 months to work through its backlog of orthopedic care needs [4, 5]. The size of the backlog for large general hospitals is alarmingly high, since these

hospitals undertake a heavier load of daily practices and, at the same time, are more responsible for pandemic control.

It should be noted that continuing to delay these surgeries could result in increased disease progression and poor health outcomes for patients. Studies have shown that delays in surgical care for osteoarthritis can result in the progressive loss of mobility and health-related quality of life; in addition, patients living with knee osteoarthritis have a higher risk of death than the general population, and their level of risk increases as their walking disabilities become more severe [6, 7]. In addition to having a potential impact on morbidity or mortality, delays in care could also lead to worse patient experiences. Given that some types of "elective" procedures are more time sensitive than others, providers may need to determine how to prioritize care provided based on the urgency of the care required and whether a delay in care could lead to morbidity or mortality [8]. In addition, the deferment of medical care has a broader impact on the national economy. It has been reported that approximately half of the annualized 4.8% United States GDP that declined in the first quarter of 2020 was attributed to health care services, especially delayed elective procedures [2].

Given the importance of surgical care on the health of patients and the financial health of hospitals, solutions to eliminate such large backlogs while maintaining the regular throughput of surgical cases are crucial to the operation of health care systems [9]. Since building additional operating rooms may not be practical for most systems given the capital costs, physical space, and associated workforce needed to do so, operators will need a robust strategic plan to effectively schedule the cumulative backlog of elective surgeries and improve patient throughput. This paper considers optimizing the use of the existing capacity of large general hospitals to work through the backlog of elective surgeries without compromising patient outcomes.

In this paper, a planning method based on the operations management framework to help large general hospitals rapidly and efficiently recover their surgical services is provided, and a scheduling algorithm for clearing elective surgery backlogs during the outbreak of a major epidemic is proposed. This method considers the balance between surgery backlogs and newly arrived patients who need elective surgeries based on waiting time and clinical urgency. It reschedules the surgeries in the backlog queue and schedules the newly arrived surgeries to ensure that all surgical services can be finished before their due day. This paper uses Markov decision process (MDP) to model the recovery process for surgeries that have been delayed due to epidemic outbreaks and designs a surgery queueing and planning algorithm based on piecewise decay ϵ-greedy reinforcement learning (PDGRL). The proposed method can dynamically determine the daily optimal plan for managing the backlog of patients awaiting surgery in the wake of a

pandemic and provide a system-based solution for delivering timely surgical care to patients and preparing for future pandemic waves.

The remainder of this paper is organized as follows. In Section 2, the relevant literature is reviewed. Section 3 describes the queueing network system for modeling the surgery backlog and recovery process. Section 4 presents the proposed stochastic scheduling optimization algorithm for surgery backlog management. Section 5 gives the simulation experimental results of implementing the proposed method on an elective surgery backlog example. Section 6 concludes the research.

2 Literature review

2.1 COVID-19 surgical backlog

Since the outbreak of COVID-19, doctors and researchers from different countries have realized and studied the consequences of surgical service delays caused by the cancellation or postponement of nonemergent examinations and treatments. Uimonen et al. [10] collected and studied the data on urologist and elective urological procedures for 2017–2020 in Finland and concluded that the health care lockdown due to COVID-19 decreased the availability of nonemergent specialized urological care; the authors also pointed out that even though these surgical services have started to recover from the delays induced by the pandemic, there is still an underlying backlog, which may result in more patients with severe conditions waiting for treatment. Fu et al. [11] presented the severe impacts of delays in surgery during the COVID-19 pandemic on patent health outcomes, hospital finances and resources, academic training and research programs, as well as health care providers. The authors suggested that the health care system should form an organized response to manage the disruption and delay of surgeries.

Many studies have used various techniques to estimate the size of backlogs, the rate of surgery rescheduling, and the reduction in outpatient and surgical activities [12], as well as the time to recovery for surgery procedures delayed due to COVID-19 [4, 13–15]. Salenger et al. [16] used simple mathematical formulations to estimate the daily backlog for cardiac surgeries during the COVID-19 pandemic and to predict the length of time required to clear this backlog. The results showed that the amount of time necessary to clear the backlog would range from 1 to 8 months based on varied estimates of postpandemic increased operational capacity and that the waiting list mortality rate could be as high as 3.7% at 1 month and 11.6% at 6 months. This study emphasized the necessity of planning for postpandemic volume and treating patients within an acceptable time

frame. Wilson et al. [17] used linear regression to quantify the volume of total hip arthroplasty and total knee arthroplasty cases delayed due to COVID-19 and to estimate the potential time required for the completion of backlogged cases. The proposed model suggested that the anticipatory planning of excess capacity and provisions for quickly and safely accommodating delayed patients are important. Wang et al. [18] used time series forecasting, queuing models and probabilistic sensitivity analysis to model the incremental surgical backlog that resulted from the COVID-19 outbreak in Ontario and to estimate the time and resources required to clear this backlog. It has been concluded that health care systems must employ innovative system-based solutions to provide patients with timely surgical care and prepare for future COVID-19 waves. Brandman et al. [19] developed an online tool to predict how long this process will take and what resources are required to address the impending backlog of elective cases based on single-entry queue models. Felfeli et al. [20] proposed microsimulation modeling using historical population-based administrative and facility-level data drawn from Ontario from January 2019 to May 2021; this model incorporated the complex dynamic navigation of patients and the interaction between resource availability and population demand to provide a more accurate estimation. The model was used to estimate the size and waiting time of the ophthalmology surgical backlogs caused by the COVID-19 pandemic and to predict the increase in provincial monthly resources required to clear these backlogs by March 2023 and by March 2022.

To clear the backlog of elective surgeries resulting from COVID-19 disruption, several authors and professional associations have discussed many clinical aspects [21] such as expanding a hospital's operating capacity by implementing enhanced recovery after surgery (ERAS) [22], maintaining capacity by adjusting the surgeon's workforce [23], adopting digitalized health care systems using telemedicine and artificial intelligence-assisted platforms [24–27], deploying technological advancements to provide frictionless and customer-centric care [28, 29], forming dedicated teams to improve operating room efficiency and thereby simplify patients' surgical care experience [2], and using surgeons' knowledge to schedule surgeries appropriately [30]. A few studies have started to explore the methods for solving the backlog problem from an operational level. Gregory et al. [31] proposed a modified cardiac-enhanced recovery program (ERP) for implementation during the COVID-19 pandemic, which tailored the standard iterative process of ERP to fit the impact of the pandemic on patients and the health care system. Martin et al. [32] constructed not only a compartmental demand model to represent various scenarios for increasing surgical capacity following the total cessation of surgical services in response to the COVID-19 pandemic but also 400 potential scenarios in which capacities were increased at different start times and rates were simulated. Valente et al. [33] applied a modified surgical waiting list infosystem (SWALIS-2020) based on the cumulative linear prioritization method to prioritize and schedule elective surgeries during COVID-19. Matava et al. [34] implemented a program based on the lean quality improvement concept to reduce the number of patients on the surgical waiting list by scheduling elective surgeries on the weekends. However, the research that addresses these surgery backlogs by optimizing the daily surgery scheduling and control and helping hospitals make operational decisions related to coping with public health emergencies is limited.

2.2 Application of reinforcement learning in health care management

To describe the stochastic and sequential decision processes, the system can be modeled by a set of system states and actions that are performed to control the system's state. The state of the system changes over time. At any time, an action can be taken that may bring rewards to the system and cause a change in the system's state. The objective is to control the system in such a way that a performance criterion is maximized. This model is called the Markov decision process (MDP); such processes are an intuitive and fundamental formalism for reinforcement learning (RL). When no prior knowledge about the MDP is presented, RL is an effective algorithm for computing good or optimal policies for problems that are modeled as an MDP [35, 36]. In the RL algorithm, an agent is connected to the system via perception and action. In each step, the agent receives information about the current state of the system, then the agent chooses an action that changes the system's state and earns rewards. The agent's behavior should aim to choose a sequence of actions that increases the long-run sum of rewards [37]. RL enables an agent to learn effective strategies in sequential decision-making problems by engaging in trial-and-error interactions with his or her environment [38].

As one of the most efficient artificial intelligence techniques, RL has undergone high levels of development in recent decades and has been applied to several health care domains. One application of the RL approach in the health care field is to support clinical decision-making, such as dynamic treatment regimes that develop a sequence of decision rules (including medication) for intervention to individualize treatments for patients based on their varying clinical characteristics and medical histories [39–46], automated clinical diagnoses that are driven by big data analysis to assist clinicians in more accurate and efficient decision-making [47–51], computer-assisted motor and cognitive rehabilitation therapies that aim to build optimal

adaptive rehabilitation strategies that are tailored to the needs of specific patients [52–54], intelligent medical imaging [55–57], health care control systems for arm motion controllers, biomedical devices, chronic disease monitoring, anesthesia controllers, and so on [40, 58–60].

Compared to clinical decision-making, the study of applying RL in health care system operations management is still in its early days. Huang et al. [61] proposed a reinforcement learning-based resource allocation mechanism and used it to optimize the resource allocation in the radiology CT-scan examination process so that the process flow time is minimized. Schutz and Kolisch [62] modeled the capacity allocation problem for multiple classes of customers and multiple types of services by a continuous-time Markov decision process which was solved by RL. They used radiology services booking system to illustrate the proposed method. Gomes [63] proposed a framework embedded with an Advantage Actor-Critic reinforcement learning algorithm for scheduling primary care doctor daily slots appointment. Lee and Lee [64] used deep RL to assign emergency department patients who need different types of treatment to available medical resources with the objective of minimizing the patient waiting time. These literatures have shown that RL is an effective way to solve dynamic and stochastic hospital operation optimization problem in which the resource allocation decisions have to be based both on the current and future states of the system. This paper extends the application of RL in postpandemic surgery recovery management based on the optimization of dynamic and stochastic scheduling.

3 Methods

3.1 Reinforcement learning and Markov decision process

RL is a learning paradigm in sequential decision making problems. It has an agent to learn how to behave in a system, where the only feedback consists of a scalar reward signal, and perform actions that maximize the reward signal in the long run. MDP is an intuitive and fundamental formalism for learning problems in stochastic domains [65]. The canonical elements of an MDP include a set of system states $\mathcal{S}$, a set of actions $\mathcal{A}$, a transition probability matrix $\mathcal{P}$ and a reward function [66]. A state $s \in \mathcal{S}$ is a unique characterization to define the model. Actions $a \in \mathcal{A}$ can be applied in some particular state to control the system. By applying action $a \in \mathcal{A}$ in a state $s \in \mathcal{S}$, the system transits from s to a new state $s' \in \mathcal{S}$, based on the transition probability distribution $P(s'|s, a)$ which is defined over possible next states. The reward function specifies rewards for being in a state $R(s)$, performing an action in a state $R(s, a)$, or transitions between states $R(s, a, s')$. The goal is to find an optimal rule of actions placed on states $\pi(s, a)$, called policy, and gather maximal rewards. To estimate how good it is for the agent to be in a certain state or to perform a certain action in a state, value function $V^\pi(s)$ is defined to represent the expected return when starting in state s and following policy π thereafter, and state-action value function $Q^\pi(s, a)$ is defined to represent the expected return starting from state s, taking action a, and thereafter following policy π.

RL is used for computing an optimal policy for an MDP with incomplete information, so that sampling and exploration are needed. The general algorithm for RL consists of interaction iterations in which the agent selects an action based on its current state, gets feedback based on the transient state and associated reward, and updates the estimated $\widetilde{V}$ and $\widetilde{Q}$. The selection of action at each step is based on the current system state s and value function V or Q. The algorithm includes exploitation and exploration mechanism to gain more reward through exploiting current knowledge around good actions, and trying out different actions to explore the system for potential better actions. Through these iterations, RL collects and updates knowledge about the system while trying and searching for better actions interactively.

3.2 Model of elective surgery backlog management system

The surgery backlog problem is modeled by a general discrete-time queueing network system M that can be formulated by a countable-state MDP and targets that use the minimum amount of time to clear the elective surgery backlog while making sure that all surgeries, including those in the backlog and those for new arrivals, are scheduled before their due days. There are 2 queues in this system. The first queue represents the patient backlog that has accumulated since the onset of a pandemic like COVID-19, which is named Q0, while the second queue represents the patients who have newly arrived at the hospital, which is named Q1. The patients in these two queues are waiting for available surgery chances, and surgeries are scheduled daily. Therefore, the time slots are positioned on typical surgical days $t = 0, 1, 2, 3, \ldots$, where $t = 1$ denotes the first day after the hospital start to resolve its surgery backlog. Initially, the patients in the first queue represent those patients whose surgeries were delayed because of the COVID-19 pandemic. Let L_0 be the number of patients in Q0 at $t = 0$. Let α^t be the total number of new patients arriving at day t. Assume that the new patients arrive in Q1 as an independent and identically distributed (i.i.d.) sequence $\{\alpha^t, t = 0, 1, \ldots\}$ with a finite rate $\lambda_1 = \mathbb{E}(\alpha^t)$. Each patient has an arrival day b_{ij}, which is defined as the day that the patient enters the queue, and a due day d_{ij}, which is defined as the day before which

the patient has to undergo surgery, where $i = 0, 1$ indicates the queue and $j = 1, 2, \ldots$ denotes the patient. The time remaining from the current day to the patient's due day represents the patient's critical level; therefore, the larger the critical level is, the less urgent the patient's surgery is. Let T be the time slot when all patients in Q0 are scheduled. Assume that the number of newly arrived patients in Q1 and the served patients at each time slot are both bounded. The system can be modeled as follows.

(1) State space $\mathcal{S}$: The system state is given by the lengths of the two queues, that is, a 2-dimensional queue backlog vector $\boldsymbol{S} = (S_0, S_1)$, $S_0 = 0, 1, \ldots, L_0$. Even though the number of newly arrived patients in Q1 at each time slot is finite, as the system changes over time, if the number of newly arrived patients in Q1 becomes larger than the number of scheduled patients in Q1 at each time slot, then the number of patients accumulated in Q1 could be infinite. Therefore, the system has an unbounded state space, which is denoted as $\mathcal{S} = \mathcal{S}_0 \times \mathcal{S}_1 = \{0, 1, \ldots, L_0\} \times \mathbb{N}$. Let S_0^t and S_1^t be the number of waiting patients in Q0 and Q1 at day t, respectively.

(2) Action space $\mathcal{A}$: The two queues compete for the services of the hospital; thus, the action is the number of patients scheduled from the two queues every day, $\boldsymbol{A} = (A_0, A_1)$. There is a maximum capacity for each surgical day, which represents the maximum number of elective surgeries that can be performed by the hospital, e.g., $L_{\max}$. Therefore, $A_o + A_1 \leq L_{\max}$, and the action space is $\mathcal{A} = \{0, 1, \ldots, \min\{L_0, L_{\max}\}\} \times \{0, 1, \ldots, L_{\max}\}$. Let $\boldsymbol{A}^t = (A_0^t, A_1^t)$ be the action taken at day t.

(3) State-transition probability P: Since the system dynamics satisfy the Markov property, the probability of the system state transitioning into a state $\boldsymbol{S'}$ only depends on the current state $\boldsymbol{S}$, the selected action $\boldsymbol{a}$ and the number of newly arrived patients, $p(\boldsymbol{S'}|\boldsymbol{S}, a)$. Each surgical day, the hospitals schedule surgeries for A_1 patients from the Q0 queue and A_2 patients from Q1, and there are α newly arrived patients that enter the queue Q1:

$$P(\boldsymbol{S'} = (s_0', s_1')|\boldsymbol{S} = (s_0, s_1), \boldsymbol{a} = (a_0, a_1))$$

$$= \begin{cases} 1 & \text{if } s_0' = s_0 - a_0 \text{ and } s_1' = s_1 - a_1 + \alpha \\ 0 & \text{otherwise} \end{cases}$$

where α follows the distribution of the number of newly arrived patients, which could be obtained from the historical surgery scheduling data or the experts' opinion.

(4) Cost $\mathcal{C}$: The objective for this problem is to schedule all the patients in Q0 as soon as possible while making sure that no patient misses his or her due day and min-

imizing the average backlog of Q1. Let x_{ij}^t be the indicator of each patient's scheduled surgery time:

$$x_{ij}^t = \begin{cases} 1 & \text{if patient } j \text{ in } Qi \text{ is scheduled at time } t \\ 0 & \text{otherwise} \end{cases}, i = 0, 1, j = 1, 2, \ldots, t = 0, 1, \ldots$$

The objectives can be summarized by the following cost function:

$$C_t = \frac{\sum_{i=0}^{1} \sum_j \left(1 - \sum_{\tau=1}^{t-1} x_{ij}^\tau\right)(t - b_{ij})}{\sum_{i=0}^{1} \sum_j \left(1 - \sum_{\tau=1}^{t-1} x_{ij}^\tau\right)}$$
$$- \frac{\sum_{i=0}^{1} \sum_j \left(1 - \sum_{\tau=1}^{t} x_{ij}^\tau\right)(d_{ij} - t)}{\sum_{i=0}^{1} \sum_j \left(1 - \sum_{\tau=1}^{t} x_{ij}^\tau\right)} \quad (1)$$
$$+ \sum_{i=0}^{1} \sum_j x_{ij}^t M\left[1 - H\left(d_{ij} - x_{ij}^t t\right)\right]$$

where $H(x)$ is the Heaviside step function whose value is zero for negative arguments and one for nonnegative arguments, $H(x) \triangleq \begin{cases} 1 & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$. At time t, an action $\boldsymbol{A}_t$ is selected, and the system has three types of patients: (1) patients scheduled at time t, (2) patients who have not been scheduled up to time t, and (3) patients who have been scheduled before time t. The first part of Eq. (1) represents the average waiting time of the patients who are scheduled at time t or have not been scheduled up to time t. The second part of Eq. (1) represents the average critical level of the patients who have not been scheduled up to time t, where the critical level is represented by the remaining number of days before the due day for each unscheduled surgery. The smaller the critical level is, the closer the surgery is to its due day and the more urgent the surgery is. The third part of Eq. (1) is the penalty term for patients who are scheduled at time t but have already missed their due day. M is a large positive penalty constant; even one overdue surgery will result in a large cost function so that the algorithm is able to prevent the occurrence of delaying a treatment. It should be noted that the cost could be negative if surgeries are scheduled long before their due days.

4 Algorithm

Since the problem has unbounded state spaces, we construct the surgery backlog scheduling algorithm based on the piecewise decaying ϵ-greedy reinforcement learning (PDGRL) algorithm proposed by Liu et al. [67] to find the optimal policy π^*. This algorithm introduces an auxiliary system $\widetilde{M}$ with bounded state space, in which each queue has buffer size U. In $\widetilde{M}$, for each queue, when the queue

backlog reaches U, new arrivals are dropped. $\widetilde{M}$ has the same action space and cost function as M. In this problem, U is selected to be greater than L_0, $U > L_0$ to ensure that all patients in Q0 will be scheduled in $\widetilde{M}$. The state space of $\widetilde{M}$ is $\widetilde{S} \triangleq \{ S \in S : S_0, S_1 \leq U, U > L_0 \}$. By exploring and exploiting the solutions within the auxiliary system while implementing the stabilizing policy outside the auxiliary system, the PDGRL algorithm provides a control policy that is close to the optimal result with a large U.

4.1 Stabilizing policy

The stabilizing policy is a policy under which the system is stable, that is, the system converges in the sense that:

$$\lim_{t \to \infty} P[S(t) \leq r] = F(r), \forall r \in S$$

where $F(\bullet)$ is the cumulative distribution function on S. According to Foss et al. [68], for a two-component Markov chain $\{(X^t, Y^t)\}$ on the state space $(\mathcal{X}, \mathcal{Y})$, where $\{X^t\}$ forms a Markov chain itself with a stationary distribution π_X, the system is stable if the following conditions hold:

A. The expectations of the absolute values of the sequence $\{L_2(Y^t)\}$ are bounded from above by a constant U:

$$\sup_{x \in \mathcal{X}, y \in \mathcal{Y}} \mathbb{E}_{x,y} \left| L_2\left(Y^{t+1}\right) - L_2\left(Y^t\right) \right| \leq U < \infty$$

where $L_2(\bullet)$ is a nonnegative measurable function that is defined on a sigma-algebra $\widetilde{\mathcal{Y}}$ of $\mathcal{Y}$.

B. There exists a nonnegative and nonincreasing function $h(N), N \geq 0$ such that $h(N) \downarrow 0$ as $N \to \infty$, and a family of mutually independent random variables $\{\varphi_x^t\}, x \in \mathcal{X}, t = 0, 1, \dots$ such that

B1. for each t, $\{\varphi_x^t\}, x \in \mathcal{X}, t = 0, 1, \dots$ are uniformly integrable;

B2. for each t and x, $\{\varphi_x^t, t = 0, 1, \dots\}$ are identically distributed with common distribution function F_x, which is such that $F_x(y)$ is measurable as a function of x for any fixed y;

B3. the following inequality:

$$L_2\left(Y^{t+1}\right) - L_2\left(Y^t\right) \leq \varphi_{X^t}^t + h\left(L_2\left(Y^t\right)\right)$$

holds for all $x \in \mathcal{X}, y \in \mathcal{Y}$ and $t = 0, 1, \dots$; and

B4. functions $f(x) = \mathbb{E}\varphi_x^1$ satisfy the following:

$$\int_{\mathcal{X}} f(x)\pi_X dx = -\varepsilon < 0.$$

Since the objective of the surgery scheduling system is to schedule all patients in both queues as soon as possible,

it is obvious that $L_{\max}$ surgeries will be scheduled every day as long as the total number of patients waiting in the system exceeds $L_{\max}$. Therefore, to simplify the proof of the stabilizing policy, the working hours on each typical surgical day t are divided into time slots $t' = L_{\max}(t-1) + 1, L_{\max}(t-1) + 2, \dots, L_{\max}t$ and $t = 1, 2, \dots$. The new patient arrival rate at t' of Q1 is $\lambda I_1 = \mathbb{E}(\alpha^{t'}) = \mathbb{E}\left(\frac{\alpha^{t'}}{L_{\max}}\right)$, where $\alpha^{t'}$ is the total number of new patients who arrive at t'. At each time t', one surgery, either from Q0 or from Q1, will be scheduled. Thus, the action at t' becomes the index of the queue being selected, $A'^{t'} \in \{0, 1\}$, and the action taken at day t is $A^t = \left(A_0^t, A_1^t\right) = \left(\sum_{t'=L_{\max}(t-1)+1}^{L_{\max}t} \left(1 - A'^{t'}\right), \sum_{t'=L_{\max}(t-1)+1}^{L_{\max}t} A'^{t'}\right)$

Consider the longest connected queue (LCQ) policy, which schedules the patient from the queue with maximum length at time slot t' as follows:

$$A'^{t'} = \begin{cases} \text{none} & \text{if } S_i^{t'} = 0, i = 0, 1 \\ \arg\max_{i=0,1} \left\{S_i^{t'}\right\} & \text{otherwise} \end{cases}$$

This is equivalent to the following:

$$A^t = \left(A_0^t, A_1^t\right) = \begin{cases} (L_{\max}, 0) \text{ if } S_1^t + L_{\max} \leq S_0^t \\ \left(\frac{L_{\max} + \left(S_0^t - S_1^t\right)}{2}, \frac{L_{\max} - \left(S_0^t - S_1^t\right)}{2}\right) \text{ if } S_1^t \leq S_0^t < S_1^t + L_{\max} \\ \left(\frac{L_{\max} - \left(S_1^t - S_0^t\right)}{2}, \frac{L_{\max} + \left(S_1^t - S_0^t\right)}{2}\right) \text{ if } S_0^t \leq S_1^t < S_0^t + L_{\max} \\ (0, L_{\max}) \text{ if } S_0^t + L_{\max} \leq S_1^t \end{cases}$$

Hence, proving the stability of $\left\{\left(S_0^t, S_1^t\right)\right\}$ is sufficient for proving the stability of $\left\{\left(S_0^{t'}, S_1^{t'}\right)\right\}$.

Theorem 1. System $\left\{\left(S_0^{t'}, S_1^{t'}\right)\right\}$ *is stable under the LCQ policy if* $\lambda I_1 = \mathbb{E}\left(\alpha^{t'}\right) \leq 1$ *and* $L_0 > 2$.

The proof of Theorem 1 is provided in the Appendix.

Theorem 1 shows that the surgery backlog system with two queues that represent the patient backlog and new arrivals, respectively, is stable under the policy of minimizing the difference between the number of unscheduled surgeries in the two queues when the mean new patient arrival rate is no more than the mean surgery process rate.

In the next section, the algorithm for surgery backlog scheduling is proposed, in which Theorem 1 is used to define the stabilizing policy as part of policy computation in each episode.

4.2 Elective surgery backlog PDGRL algorithm

The algorithm operates in an episodic manner. In each episode, it either performs exploration or exploitation. In the exploration stage, the algorithm applies a random policy π_{rand}, which selects an action in $\mathcal{A}$ uniformly, to each state

in $\tilde{\mathcal{S}}$ and applies the stabilizing policy π_{stable} to each state in $\mathcal{S}\backslash\tilde{\mathcal{S}}$. In the exploitation stage, the reward matrix R is estimated by the calculations made in the previous episodes and an MDP consisting of transition matrix P, after which an updated reward matrix R is solved, and the optimal policy π^* is obtained. The algorithm applies π^* to each state in $\mathcal{S}^{in} \triangleq \left\{ \boldsymbol{S} \in \mathcal{S} : S_0 + S_1 \leq U - L_{\max} \right\}$ and applies π_{stable} to each state in $\mathcal{S}\backslash\mathcal{S}^{in}$. The detailed algorithm is shown as follows.

Input: $\mathcal{S}$, $\mathcal{A}$, Q0 = (patient ID, b_{ij}, d_{ij}, surgery scheduled day), U, $l \in (0,1]$, K, $L_{\max}$, M, P, and K.

Initialization: R_1, in which each element corresponding to the nonzero elements in P is 100.

For episodes $k = 1,2,\ldots,K$ **do**

    Initialization of episode k:

    1. Set Q1 = (patient ID, b_{ij}, d_{ij}, surgery scheduled day) is an empty matrix.

    2. Compute $\epsilon_k = \frac{l}{\sqrt{k}}$ for defining the threshold of deciding whether to explore or exploit during episode k.

    3. Uniformly generate a random number $\xi \in [0,1]$.

    4. Set $t := 1$.

    5. Set the state-action-state pair matrix, $QAQ = (t, S_0^{t-1}, S_1^{t-1}, A_0^t, A_1^t, S_0^t, S_1^t)$, in which each row stores the current time, the state of the last time, the action taken and the state of the current time, to be $QAQ := (0, \text{Q0 length}, 0, 0, 0, 0, 0)$.

    6. Set $\text{Cost}_k := 0$.

    Compute policy π_k of episode k:

    If $\xi \leq \epsilon_k$ **do**

    7. Set $\pi_k(\boldsymbol{S}) = \begin{cases} \pi_{\text{rand}}(\boldsymbol{S}), \text{for } \boldsymbol{S} \in \tilde{\mathcal{S}} \\ \pi_{\text{stable}}(\boldsymbol{S}), \text{for } \boldsymbol{S} \in \mathcal{S}\backslash\tilde{\mathcal{S}} \end{cases}.$

    else do

    8. Solve the MDP of P and R_k, and obtain the optimal policy π_k^*.

$$9. \text{ Set } \pi_k(\boldsymbol{S}) = \begin{cases} \pi_k^*(\boldsymbol{S}), \text{for } \boldsymbol{S} \in \mathcal{S}^{in} \\ \pi_{\text{stable}}(\boldsymbol{S}), \text{for } \boldsymbol{S} \in \mathcal{S} \backslash \mathcal{S}^{in} \end{cases}.$$

End if

Update R_k of episode k:

While Q0 is not empty **do**

    10. Take action $(a_0^t, a_1^t) = \pi_k(s_0^{t-1}, s_1^{t-1})$.

    11. Implement (a_0^t, a_1^t) into the system and observe the next state (s_0^t, s_1^t).

    12. Update QAQ by $QAQ := \begin{pmatrix} QAQ \\ t, s_0^{t-1}, s_1^{t-1}, a_0^t, a_1^t, s_0^t, s_1^t \end{pmatrix}$.

    13. Update Q0 by removing a_0^t rows, and update Q1 by removing a_1^t rows.

    14. Generate $s_1^t - (s_1^{t-1} - a_1^t)$ newly arrived patients for Q1, and update Q1 by adding $s_1^t - (s_1^{t-1} - a_1^t)$ rows.

    15. Set $t := t + 1$.

End while

For rows in QAQ $p = 2, \dots,$ number of rows of QAQ **do**

    16. Take $t, \boldsymbol{S} = (s_0^{t-1}, s_1^{t-1})$, $\boldsymbol{A} = (a_0^t, a_1^t)$, and $\boldsymbol{S}' = (s_0^t, s_1^t)$.

    17. Calculate the reward Cost_p by Eq. (1).

    18. Set $R_k(\boldsymbol{S}, \boldsymbol{S}', \boldsymbol{A}) = \text{Cost}_p$.

    19. Set $\text{Cost}_k := \text{Cost}_k + \text{Cost}_p$.

End for

End for

Output: π_K.

5 Computational experiments

5.1 Experiments on simulated datasets

The model and algorithm proposed in the previous sections are first tested on a series of simulated datasets, to verify its effectiveness and study the robustness of its performance. The main parameters of the problem are as follows:

1. Total number of backlogged elective surgeries by the time when hospital starts to recover its surgical services, L_0: The larger the hospital is, or the longer the peak weeks in one wave of pandemic, the larger L_0 is. A larger L_0 means that it will take a longer time to clear the backlog, and the probability of patients missing their due days will be higher.

2. Number of elective surgeries that can be done every day from the hospital starts to recover its surgical services until the hospital fully recovers its capacity, $L_{\max}$: The larger $L_{\max}$ is, the faster the backlog will be cleared.

3. Proportions of patients with different critical levels. In this research, based on the levels of clinical severity, the surgeries are classified into three types: due in 30 days,

Table 1 Simulation experiments setup

Experiment number	Total number of backlogged surgeries	Service capacity (per day)	Proportions of patients with different critical levels (due in 30 days, due in 60 days, due in 90 days)	Earliest original scheduled day of awaiting backlogged surgeries	Mean of newly arrived patients (per day)
1	10,000	100	(30%, 40%, 30%)	-30	80
2	10,000	100	(10%, 30%, 60%)	-15	80
3	6000	100	(30%, 40%, 30%)	-30	80
4	6000	100	(10%, 30%, 60%)	-15	80
5	10,000	100	(30%, 40%, 30%)	-30	120
6	10,000	100	(10%, 30%, 60%)	-15	120
7	6000	100	(30%, 40%, 30%)	-30	120
8	6000	100	(10%, 30%, 60%)	-15	120

due in 60 days, and due in 90 days. If there are a large proportion of surgeries with high critical level being delayed during the pandemic, the system may need to schedule more surgeries from the backlog queue to reduce the risk of increased disease progression and poor health outcomes. However, the chance of patients missing their due days will still be higher if the proportion of patients with severe illness is increased.

4. Earliest original scheduled day of awaiting patients in the backlog queue: If the hospital stops or cuts its elective surgical services for a long time, the number of patients who have been waiting in the backlog queue passing their due days will be large. Even some patients with low critical levels will be close to their due days at the beginning of the hospital's postpandemic recovery. Thus, there will be many patients who need to be scheduled immediately after the recovery plan starts, which may exceed the service capacity, and cause that more patients cannot be scheduled before their due days.

5. Mean of newly arrived patients every day after the hospital starts to recover its surgical services: The proposed algorithm is designed to clear the backlog as soon as possible, while minimizing the number of surgeries scheduled after the due day. This procedure considers the critical level of surgeries both in the backlog queue and in the new arrivals queue. Therefore, the rate of new arrivals in queue Q1 will affect the time needed to clear the backlog. If the rate of new arrivals is extremely high, more patients with high critical levels need to be included in the schedule and compete for the limited service capacity with the backlogged patients.

Table 1 summarizes the setup of simulation experiments for evaluating the performance of the proposed method.

Each setup was run for 20 replications, and each replication was run for 1000 episodes to test the convergence. The statistics of the number of days to clear the backlog, the number of patients who miss their due day, and running time are summarized in Table 2.

The experiments were conducted with $U = 50, l = 0.5$, and $M = 1000$. These simulation setups include some extreme cases. Take the first experiment as an example. Compared to the number of surgeries that can be done each day, the size of the backlog is very large. Meanwhile, the

Table 2 Comparison of simulation experiments

Experiment number	95% confidence interval of average number of days to clear the backlog (days)	95% confidence interval of average number of patients who miss their due day	95% confidence interval of average running time (seconds)
1	[124.42, 144.68]	[12136.81, 12,223.79]	[90820.32, 129,151.40]
2	[130.13, 136.17]	[77.67, 158.63]	[10092.77, 12,903.41]
3	[71.43, 80.67]	[11.29,32.71]	[17807.91, 34,905.78]
4	[62.18, 65.02]	[0.07, 1.55]	[18180.63, 34,349.12]
5	[140.21, 165.99]	[13342.81, 13,346.69]	[114977.29, [169723.80]
6	[138.18, 160.12]	[75.92, 204.14]	[122783.10, 17,188.43]
7	[77.37, 89.33]	[20.96, 58.30]	[17688.81, 34,637.51]
8	[65.46, 74.64]	[0.04, 3.71]	[17457.56, 25,682.46]

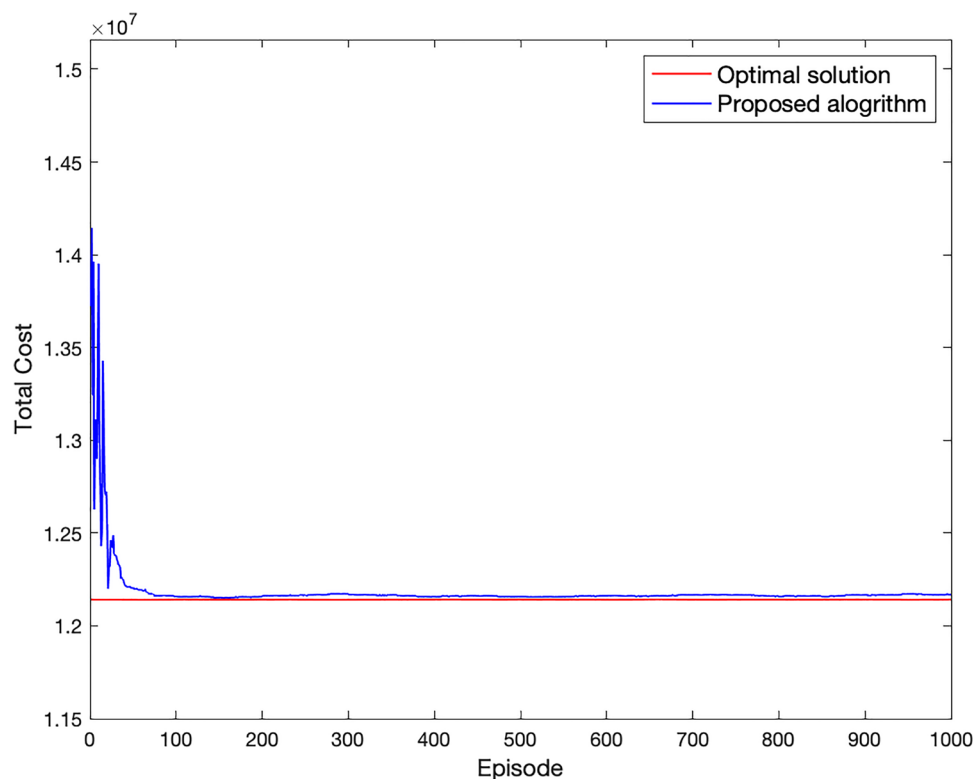**Fig. 1** Episode cost function of simulation experiment 1 in Table 1

Table 3 Information of surgical backlog (Q0)

Surgery number	Arrival day	Due day	Original scheduled day
1	-4	86	28
2	-57	33	26
3	-33	27	-19
4	-11	49	16
5	-7	83	32
⋮	⋮	⋮	⋮

proportions of surgeries with high critical levels are very large too. Therefore, in the simulated $Q0$ queue, there are more surgeries due on each day than the daily maximum surgical service capacity. For such cases, the optimal scheduling policy would be to schedule all surgeries in both $Q0$ queue and $Q1$ queue based on their due days, even though the cases of missing due days would happen from the first day to the last day when all surgeries in $Q0$ queue have been cleared. Figure 1 gives the episode cost function of the first experiment to illustrate the convergence of the algorithm.

Based on the simulation experiments, it can be seen that the proposed elective surgeries backlog PDGRL algorithm is able to solve different problems with various settings of parameters, and can converge to the optimal solution within $80 \sim 150$ episodes. For extreme cases in which the size of the

backlog is relatively large compared with the daily capacity, and the proportions of high critical level patients are large, the algorithm can give solutions close to the optimal scheduling. For normal cases, the algorithm can give solutions with very few surgeries missing the due day. However, the algorithm running time is sensitive to the size of the problem. As the size of the backlog and daily service capacity increase, the computation speed decreases significantly because of the slower MDP solving for large state space and action space.

5.2 Numerical example

W Hospital, which is located in Western China, is one of the largest hospitals nationwide. During the peak weeks of the COVID-19 outbreak in China, elective surgery operations in W Hospital were delayed from January 27 to March 27, 2020, and a total of 16,377 elective surgeries were awaiting by the end of March, which is when the hospital started to use its medical partnerships to recover its surgical services. Let $t = 0$ be the start time when the hospital started to clear the surgical backlog. Thus, $t = -1. -2, \ldots$ denote one day, two days, and so on before the start time, and $t = 1, 2, \ldots$ denote one day, two days, and so on after the start time. The information pertaining to each surgery includes the arrival day, due day, and original scheduled day, as shown in Table 3. While the surgical backlog was cleared, there were newly arrived patients who needed elective surgeries

Table 4 Comparison of different surgery backlog clearing strategies

	Real data	Real ad hoc policy used by W Hospital	First-come, first-served (FCFS)	Earliest due date (EDD)	Best policy obtained by elective surgeries backlog PDGRL algorithm with priority placed on following the original scheduled arrangement	Best policy obtained by the elective surgeries backlog PDGRL algorithm with priority placed on due days
Total cost	3,279,066.14	4,072,710.73	3,694,767.41	572.91	9044.91	-1319.46
Number of days to clear the backlog	56	42	42	60	47	49
Number of patients missing their due days (including surgeries both in $Q0$ and $Q1$	3288	4071	3693	0	10	0

every day. Based on the historical data, the number of newly arrived patients every day follows a Poisson distribution with a mean of 300, in which 19.3%, 39.1%, and 41.6% of these patients needed their surgeries to be conducted within 30 days, 60 days, and 90 days, respectively.

The 16,377 elective surgeries form queue Q0, and the newly arrived surgeries form queue Q1. The algorithm proposed in the previous section is applied to reschedule the surgeries in Q0 and to schedule the newly arrived surgeries in Q1; the overall objective is to clear the Q0 backlog as soon as possible while ensuring that the number of patients who miss their due day is minimized. The capacities of W Hospital and its branch health center locations determine that the number of elective surgeries that can be performed each day is 390 on average.

The benchmark policies to compare with are the first-come first-served (FCFS), the earliest due date (EDD), and the real policy that had been implemented in W Hospital during their surgical service recovery after the peak weeks of the COVID-19 pandemic, which was done ad hoc. The FCFS policy schedules the surgeries according to the sequence patients arrive. The EDD policy schedules the surgeries according to the due dates of surgeries. The real ad hoc policy used by W hospital scheduled those backlogged surgeries first based on the original scheduled arrangement, and didn't start to schedule newly arrived patients until all backlogged surgeries have been cleared. Approximately 9.72% of the backlogged patients was worsened precipitously, so that they left the waiting queue and took emergency operations.

First, the proposed method is implemented with a priority placed on following the original scheduled arrangement. Second, the proposed method is implemented with a priority placed on the due days, that is, the surgeries that are closer to their due days are scheduled first. The algorithm was conducted with $U = 50, l = 0.5, M = 1000$, and $K = 200$. Table 4 shows the comparison of real data,

real ad hoc policy used by W Hospital, FCFS policy, EDD policy, the optimal policy obtained by the elective surgeries backlog PDGRL algorithm with a priority placed on following the original scheduled arrangement, and the optimal policy obtained by the elective surgeries backlog PDGRL algorithm with a priority placed on due days.

The real data was collected from March 30, 2020 to May 25, 2020 when the backlog had been cleared, to show the real situation that happened after the pandemic outbreak. According to the head nurse who was in charge of elective surgery scheduling, the main reasons for the difference between real data and read ad hoc policy used are: (1) Some doctors were not available at the beginning of the service recovery phase, which delayed the surgeries they were in charge of. (2) Some patients left the waiting list. They might go to hospitals near their place of residence to receive faster treatments. This caused the number of patients in the $Q0$ queue to be less than 16,377.

From Table 4, it can be seen that the optimal policy obtained by the proposed elective surgeries backlog PDGRL algorithm with a priority placed on due days could clear the backlogged surgeries in a shorter time while minimizing the number of patients missing their due days.

Unlike Q0, in which all the surgeries waiting are known when the hospitals start to clear their backlog and no newly arrived surgeries are entering this queue, the surgeries in Q1 are newly arrived and unknown at the beginning of the clearing process. Therefore, scheduling decisions are dynamically made every day based on the surgeries remaining unscheduled in Q0 and the number of surgeries waiting in Q1 on the previous day. To examine the performance of the proposed method, 100 Q1 cases are simulated, and the obtained optimal strategy is implemented for each case. Based on the comparison shown in the Table 4, the elective surgeries backlog PDGRL algorithm with a priority placed on due days
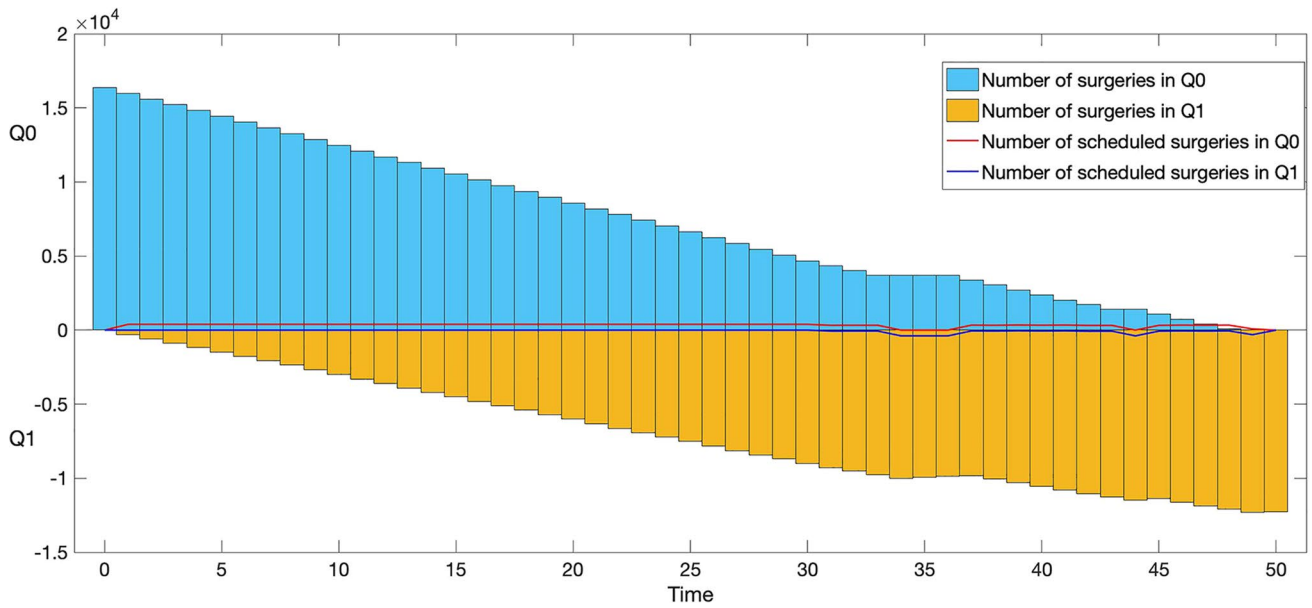
Fig. 2 Optimal policy: number of surgeries scheduled every day for Q0 and Q1 from the randomly selected cases

for the randomly selected cases outperforms the algorithm with a priority placed on following the original scheduled arrangement. Figures 2 and 3 show the optimal scheduling strategy obtained by the elective surgeries backlog PDGRL algorithm with a priority placed on due days for the randomly selected cases.
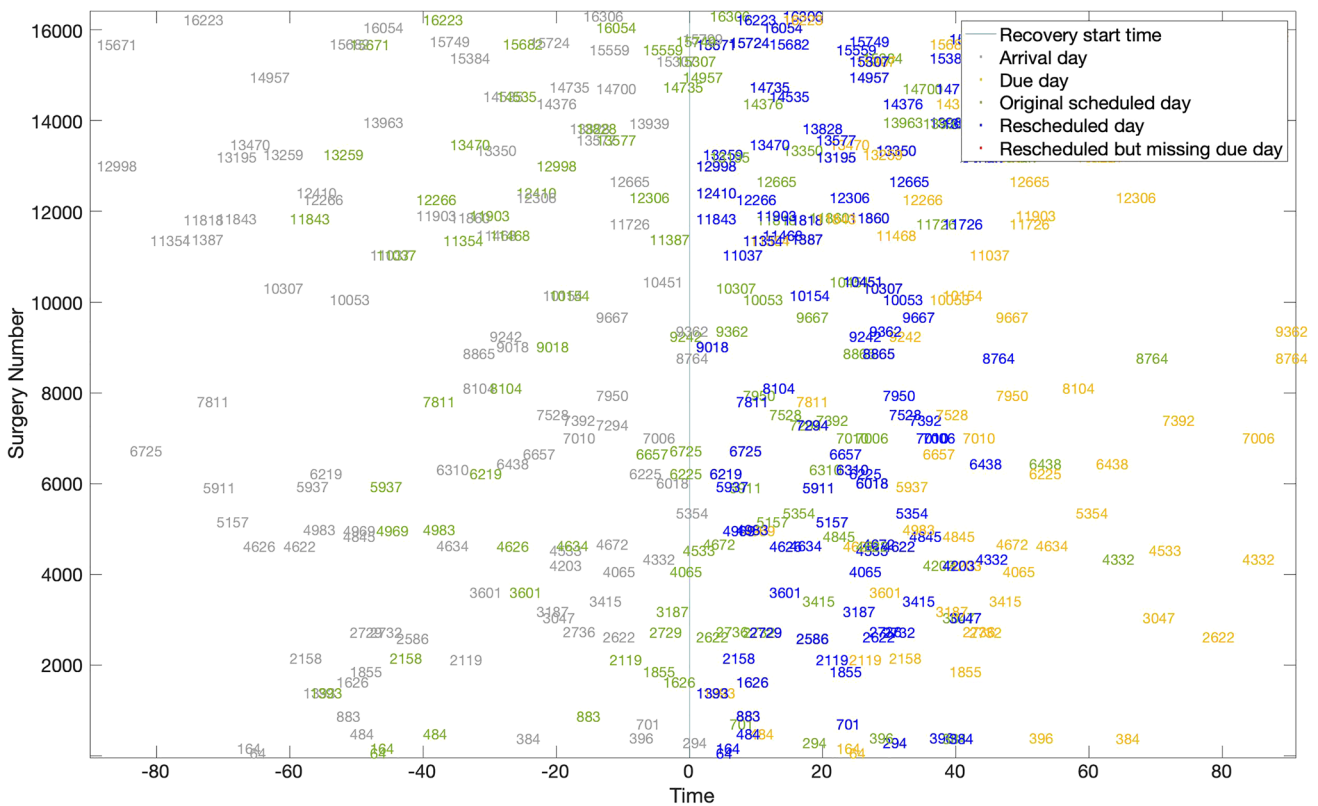

Fig. 3 Optimal policy: timeline of 100 randomly selected surgeries from Q0

Fig. 4 Histogram of (**a**) number of days to clear the backlog $Q0$, and (**b**) number of surgeries missing due days for the optimal policy with 100 $Q1$ simulated cases
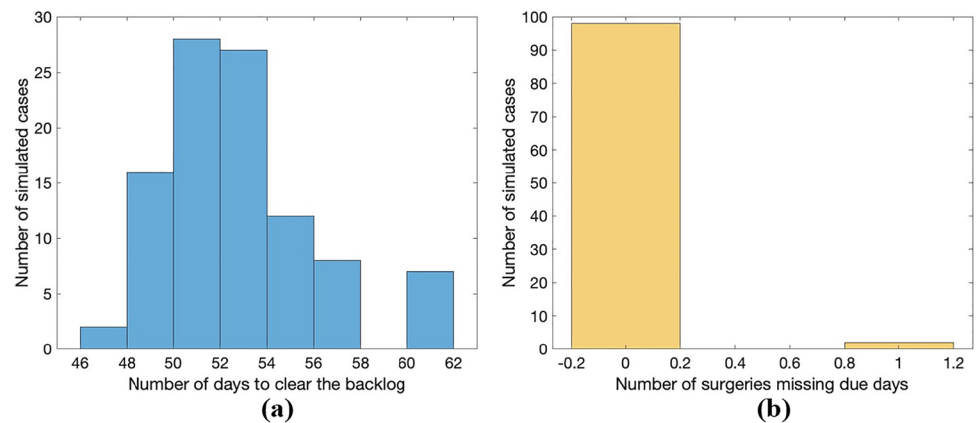

(a)

(b)

Figure 2 plots the number of surgeries scheduled every day for Q0 and Q1. Figure 3 plots the timeline of the 100 randomly selected surgeries from Q0, where the arrival day of each surgery is marked by a gray number, the due day of each surgery is marked by an orange number, the original scheduled day of each surgery is marked by a green number, and the rescheduled day of each surgery is marked by a blue number. If there is any surgery that has missed its due day, its rescheduled day is marked in red. By summarizing the 100 simulated cases, it is found that the mean time that it will take W Hospital to clear its elective surgery backlog is 51.72 days (95% confidence interval: (51.01, 53.47)), and the average number of surgeries (both in Q0 and Q1) that are not scheduled before their due day is 0.02 (95% interval: (0, 0.048)). The histograms of performance measures of the 100 simulated cases are shown in Fig. 4.

5.3 Summary and discussion

Through simulation experiments and real data numerical example, the proposed elective surgeries backlog PDGRL algorithm has shown good potential of tackling surgery backlog caused by a pandemic such as COVID-19. In the simulation study, extreme cases in which the number of backlogged surgeries closing to due days is too large was tested, and the results show that the proposed algorithm can quickly converge to the optimal solution. The performance of the algorithm is not sensitive to the parameters of the problem, except that the running time is greatly affected by the size of the problem.

Compared to the ad hoc procedure which was used in the hospital, the proposed elective surgeries backlog PDGRL algorithm could improve the after-pandemic surgery backlog scheduling performance by reducing the average waiting time of patients as well as the risk of missing the best treatment time. The ad hoc procedures are designed to schedule the backlog first following the original scheduled arrangement, without considering the newly arrived patients and the critical levels of patients. This leads to a short backlog clearance time, but a large number of patients who miss their best treatment opportunity, and an increasing waiting time for all patients in the system. The proposed algorithm could find an optimal balance between these factors. However, the difference between the calculated policy with the real data indicates that more realistic factors such as the patients' behavior, the after-pandemic recovery plans for different hospitals, and the doctors' preference, need to be included in the algorithm in the future to create an accurate surgery backlog management system.

The most serious problem with the proposed elective surgeries backlog PDGRL algorithm is that its computation speed is highly affected by the size of the backlog and the maximum number of surgeries that can be done every day, since the increase in these two parameters will significantly increase the volume of state space $\mathcal{S}$ and action space $\mathcal{A}$, which in turn makes the search space exponentially growing. For a problem with 10,000 backlogged surgeries to be cleared, it may take more than a day to run the algorithm for 1000 episodes, compared to five hours for the problem with 6000 backlogged surgeries. This would limit the application of this algorithm when fast decisions are required, which needs future work to improve the computation speed.

6 Conclusion

A large number of elective surgeries are postponed due to disruptions caused by pandemics, such as COVID-19, which results in serious backlogs. Continuing to delay these surgeries could result in disease progression and poor health outcomes for patients, as well as financial losses for health care systems and nations. This paper presented a stochastic control process-based method that aims to help large hospitals make operational recovery plans to clear their elective surgery backlog and restore surgical activity safely. The proposed solution uses the MDP model and reinforcement learning algorithm and shows

effectiveness in managing elective surgeries during pandemics. It can be adapted to a hospital's decision support system using local data to assist with health care system recovery planning and help hospitals prepare for future pandemic waves. For future work, the algorithm could be extended to allow dynamic changes in hospital capacities so that the stochastic scheduling optimization can realize the real-time management of elective surgeries during public health emergencies, include more factors in the algorithm such as doctors' preferences and surgery due date windows to make the system more accurate and practical, and consider the patients' choices in case that some patients are not available on their scheduled days.

Appendix

Proof of Theorem 1

Proof. The sequence $\left\{ S_1^{t'}, t' = 0, 1, \ldots \right\}$ forms a Markov chain with

$$P\left(s_1^{t'+1} | s_1^{t'}, a'^{t'} \right) = \begin{cases} 1 & \text{if } s_1^{t'+1} = s_1^{t'} + \alpha^{t'} - 1 \text{ and } a'^{t'} = 1 \\ 1 & \text{if } s_1^{t'+1} = s_1^{t'} + \alpha^{t'} \text{ and } a'^{t'} = 0 \\ 0 & \text{otherwise} \end{cases}$$

and a stationary distribution $\pi_{S_1} = P\left(S_1^{t'} = \lceil \frac{L_{\max}}{2} \rceil \right) = 1$. Take function $L_2(y) = y$, and consider the state of the Markov chain $\left\{ \left(S_0^{t'}, S_1^{t'} \right) \right\}$ for $t' = 0, 1, \ldots,$

$$\left| L_2\left(S_0^{t'+1} \right) - L_2\left(S_0^{t'} \right) \right| = \left| S_0^{t'+1} - S_0^{t'} \right| = \left| \left[L_0 - \sum_{\tau'=1}^{t'+1} \left(1 - A'^{\tau'} \right) \right] - \left[L_0 - \sum_{\tau'=1}^{t'} \left(1 - A'^{\tau'} \right) \right] \right| = \left| -\left(1 - A'^{t'+1} \right) \right| = 1 - A'^{t'+1}$$

$$\mathbb{E}_{S_0, S_1} \left| L_2\left(S_0^{t'+1} \right) - L_2\left(S_0^{t'} \right) \right| = \mathbb{E}_{S_0, S_1} \left(1 - A'^{t'+1} \right) = P\left(S_0^{t'+1} > S_1^{t'+1} \right)(1 - 0) + P\left(S_0^{t'+1} < S_1^{t'+1} \right)(1 - 1) = P\left(S_0^{t'+1} > S_1^{t'+1} \right) \leq 1 < \infty$$

Thus, Condition A holds.

$$\varphi_{s_1}^{t'} = S_1^{t'} - \left\{ L_0 - \mathbb{I}(L_0 > \alpha^1) - \mathbb{I}\left(L_0 > \sum_{\tau'=1}^{2} \alpha^{\tau'} - 1 + 2\mathbb{I}(L_0 > \alpha^1) \right) - \right.$$

$$\mathbb{I}\left(L_0 > \sum_{\tau'=1}^{3} \alpha^{\tau'} - 2 + 2\mathbb{I}(L_0 > \alpha^1) + 2\mathbb{I}\left(L_0 > \sum_{\tau'=1}^{2} \alpha^{\tau'} - 1 + 2\mathbb{I}(L_0 > \alpha^1) \right) \right) - \cdots$$

$$\left. + 2\mathbb{I}(L_0 > \alpha^1) + \ldots 2\mathbb{I}\left(L_0 > \sum_{\tau'=1}^{t'-2} \alpha^{\tau'} - t' + 3 + 2\mathbb{I}(L_0 > \alpha^1) + \ldots \right) \right) \right) \right\}, \text{ where } \mathbb{I}(\bullet)$$

is the indicator function $\mathbb{I}(\bullet) = \begin{cases} 1 \text{ if } \bullet \text{ is true} \\ 0 \text{ if } \bullet \text{ is false} \end{cases}$. For each t', take $K = 0$, and under the LCQ policy, we have the following:

$$\mathbb{E}\left[\left| \varphi_{s_1}^{t'} \right| \mathbb{I}\left(\left| \varphi_{s_1}^{t'} \right| \geq K \right) \right]$$

$$\leq \mathbb{E}\left[\left| \varphi_{s_1}^{t'} \right| \right]$$

$$= \mathbb{E}\left[\left| S_1^{t'} - L_0 + \mathbb{I}(L_0 > \alpha^1) + \mathbb{I}\left(L_0 > \sum_{\tau'=1}^{2} \alpha^{\tau'} - 1 + 2\mathbb{I}(L_0 > \alpha^1) \right) \right. \right.$$

$$\left. + \mathbb{I}\left(L_0 > \sum_{\tau'=1}^{3} \alpha^{\tau'} - 2 + 2\mathbb{I}(L_0 > \alpha^1) + 2\mathbb{I}\left(L_0 > \sum_{\tau'=1}^{2} \alpha^{\tau'} - 1 + 2\mathbb{I}(L_0 > \alpha^1) \right) \right) \right)$$

$$\left. + \cdots + \mathbb{I}\left(L_0 > \sum_{\tau'=1}^{t'} \alpha^{\tau'} - t' + 1 + 2\mathbb{I}(L_0 > \alpha^1) + \cdots + 2\mathbb{I}\left(L_0 > \sum_{\tau'=1}^{t'-1} \alpha^{\tau'} - t' + 2 + \right. \right. \right.$$

$$\left. \left. \left. 2\mathbb{I}(L_0 > \alpha^1) + \ldots 2\mathbb{I}\left(L_0 > \sum_{\tau'=1}^{t'-2} \alpha^{\tau'} - t' + 3 + 2\mathbb{I}(L_0 > \alpha^1) + \ldots \right) \right) \right) \right) \right| \right]$$

$$\leq \mathbb{E}\left[\left| S_1^{t'} - L_0 \right| \right]$$

$$\leq \frac{L_0}{2}$$

if $\lambda'_1 = \mathbb{E}\left(\alpha^{t'} \right) \leq 1$. Therefore, $\left\{ \varphi_{s_1}^{t'} \right\}$ are uniformly integrable, and assumption B1 holds. Since $\varphi_{s_1}^{t'}$ is the difference between the length of Q0 and Q1 at each time slot t', it is obvious that they are i.i.d. under the LCQ policy, which reduces the length of the longer queue by 1. For any fixed $s_0^{t'}$, $S_1^{t'+1} = S_1^{t'} - 1$ with probability $P\left(S_1^{t'} \geq s_0^{t'} \right)$, and $S_1^{t'+1} = S_1^{t'}$ with probability $P\left(S_1^{t'} < s_0^{t'} \right)$. Thus, assumption B2 holds. Let $h(y) = L_0 \mathbb{I}\left(y \leq L_0 \right)$. Since

$$L_2\left(S_0^{t'+1} \right) - L_2\left(S_0^{t'} \right) = -\left(1 - A'^{t'+1} \right) \leq 0$$

and for each $t' = 0, 1, 2, \ldots$

$$\varphi_{s_1}^{t'} + h(S_0^{t'})$$

$$= S_1^{t'} - L_0 + \mathbb{I}(L_0 > \alpha^1) + \mathbb{I}\left(L_0 > \sum_{\tau'=1}^{2} \alpha^{\tau'} - 1 + 2\mathbb{I}(L_0 > \alpha^1) \right)$$

$$+ \mathbb{I}\left(L_0 > \sum_{\tau'=1}^{3} \alpha^{\tau'} - 2 + 2\mathbb{I}(L_0 > \alpha^1) + 2\mathbb{I}\left(L_0 > \sum_{\tau'=1}^{2} \alpha^{\tau'} - 1 + 2\mathbb{I}(L_0 > \alpha^1) \right) \right)$$

$$+ \cdots + \mathbb{I}\left(L_0 > \sum_{\tau'=1}^{t'} \alpha^{\tau'} - t' + 1 + 2\mathbb{I}(L_0 > \alpha^1) + \cdots + 2\mathbb{I}\left(L_0 > \sum_{\tau'=1}^{t'-1} \alpha^{\tau'} - t' + 2 + \right. \right.$$

$$\left. \left. 2\mathbb{I}(L_0 > \alpha^1) + \ldots 2\mathbb{I}\left(L_0 > \sum_{\tau'=1}^{t'-2} \alpha^{\tau'} - t' + 3 + 2\mathbb{I}(L_0 > \alpha^1) + \ldots \right) \right) \right) + L_0 \mathbb{I}(S_0^{t'} \leq L_0)$$

$$\geq S_1^{t'} - L_0 + L_0 \mathbb{I}(S_0^{t'} \leq L_0)$$

$$= S_1^{t'} - L_0 + L_0$$

$$= S_1^{t'}$$

$$\geq 0$$

we have $L_2\left(S_0^{t'+1} \right) - L_2\left(S_0^{t'} \right) \leq \varphi_{s_1}^{t'} + h\left(S_0^{t'} \right)$, and assumption B3 holds. In addition,

$$\mathbb{E}\left[\varphi_{s_1}^1 \right]$$

$$= \mathbb{E}\left[S_1^{t'} - L_0 + \mathbb{I}\left(L_0 > \alpha^1 \right) \right]$$

$$= \mathbb{E}\left[S_1^{t'} \right] - L_0 + \left[\mathbb{I}\left(L_0 > \alpha^1 \right) \right]$$

$$= \mathbb{E}\left[S_1^{t'} \right] - L_0 + P\left(L_0 > \alpha^1 \right)$$

$$\leq 1 - L_0 + 1$$

$$= 2 - L_0$$

$$< 0$$

if $L_0 > 2$. Therefore, $\int_{\chi} \mathbb{E}\left[\varphi_{s_1}^1 \right] \pi_{S_1} dx < 0$, and assumption B4 holds. Therefore, if $\lambda'_1 = \mathbb{E}\left(\alpha^{t'} \right) \leq 1$ and $L_0 > 2$, then LCQ is the stabilizing policy of the system $\left\{ \left(S_0^{t'}, S_1^{t'} \right) \right\}$. ∎

Declarations

Conflicts of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Urban K (2020) How to Improve the Surgery Backlog During COVID-19. Michigan Medicine Health Lab. https://labblog.uofmhealth.org/lab-notes/how-to-improve-surgery-backlog-during-covid-19. Accessed 2 May 2021

2. Jain A, Dai T, Bibee K, Myers CG (2020) Covid-19 Created an Elective Surgery Backlog. How Can Hospitals Get Back on Track? Harvard Business Review. https://hbr.org/2020/08/covid-19-created-an-elective-surgery-backlog-how-can-hospitals-get-back-on-track. Accessed 2 May 2021

3. Negopdiev D, Collaborative C, Hoste E (2020) Elective surgery cancellations due to the COVID-19 pandemic: global predictive modelling to inform surgical recovery plans. Br J Surg 107(11):1440–1449. https://doi.org/10.1002/bjs.11746

4. Jain A, Jain P, Aggarwal S (2020) SARS-CoV-2 impact on elective orthopaedic surgery: implications for post-pandemic recovery. J Bone Joint Surg Am 102(13):e68. https://doi.org/10.2106/JBJS.20.00602

5. Aggarwal S, Jain P, Jain A (2020) COVID-19 and cataract surgery backlog in Medicare beneficiaries. J Cataract Refract Surg 46(11):1530–1533. https://doi.org/10.1097/j.jcrs.0000000000000337

6. Hart A, Bergeron SG, Epure L, Huk O, Zukor D, Antoniou J (2015) Comparison of US and Canadian perioperative outcomes and hospital efficiency after total hip and knee arthroplasty. JAMA Surg 150(10):990–998. https://doi.org/10.1001/jamasurg.2015.1239

7. Nüesch E, Dieppe P, Reichenbach S, Williams S, Iff S, Jüni P (2011) All cause and disease specific mortality in patients with knee or hip osteoarthritis: population based cohort study. BMJ 342:d1165. https://doi.org/10.1136/bmj.d1165

8. Berlin G, Bueno D, Gibler K, Schulz J (2020) Cutting Through the COVID-19 Surgical Backlog. McKinsey & Company. https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/cutting-through-the-covid-19-surgical-backlog#. Accessed 2 May 2021

9. Samson B (2021) The difficult clearance of the elective surgical backlog caused by the cancellation of cases due to the COVID-19 pandemic. Can J Anesth/J Can Anesth 68:932–933. https://doi.org/10.1007/s12630-021-01952-0

10. Uimonen M, Kuitunen I, Seikkula H, Mattila V, Ponkilainen V (2021) Healthcare lockdown resulted in a treatment backlog in elective urological surgery during COVID-19. BJU Int 128(1):33–35

11. Fu S, George E, Maggio P, Hawn M, Nazerali R (2020) The consequences of delaying elective surgery: surgical perspective. Ann Surg 272(2):e79–e80. https://doi.org/10.1097/SLA.0000000000003998

12. Ting D, Deshmukh R, Said D, Dua H (2020) The impact of COVID-19 pandemic on ophthalmology services: are we ready for the aftermath? Ther Adv Ophthalmol 12:2515841420964099. https://doi.org/10.1177/2515841420964099

13. Anderson EG, Freeman R, Joglekar N (2020) Ramping up elective surgery after COVID-19 disruption: service capacity analysis. SSRN. https://ssrn.com/abstract=3606761. Accessed 2 May 2021

14. Brown CS, Albright J, Henke PK, Mansour MA, Weaver M, Osborne NH (2021) Modeling the elective vascular surgery recovery after coronavirus disease 2019: implications for moving forward. J Vasc Surg 73(6):1876–1880. https://doi.org/10.1016/j.jvs.2020.11.025

15. Sagalow E, Duffy A, Selvakumar P, Cognetti E (2021) Otolaryngology subspecialty surgical case rescheduling rate during the COVID-19 pandemic. Otolaryngol-Head Neck Surg 165(1_suppl):285–286. https://doi.org/10.1177/01945998211030910f

16. Salenger R, Etchill E, Ad N, Matthew T, Alejo D, Whitman G, Lawton J, Lau C, Gammie C, Gammie J (2020) The surge after the surge: cardiac surgery post-COVID-19. Ann Thorac Surg 110(6):2020–2025. https://doi.org/10.1016/j.athoracsur.2020.04.018

17. Wilson JM, Schwartz AM, Farley KX, Roberson JR, Bradbury TL, Guild GN (2020) Quantifying the backlog of total hip and knee arthroplasty cases: predicting the impact of COVID-19. HSS J 16(1_suppl):85–91. https://doi.org/10.1007/s11420-020-09806-z

18. Wang J, Vahid S, Eberg M, Milroy S, Milkovich J, Wright FC, Hunter A, Kalladeen R, Zanchetta C, Wijeysundera HC, Irish J (2020) Clearing the surgical backlog caused by COVID-19 in Ontario: a time series modelling study. CMAJ 192(44):E1347–E1356. https://doi.org/10.1503/cmaj.201521

19. Brandman DM, Leck E, Christie S (2020) Modelling the backlog of COVID-19 cases for a surgical group. Can J Surg 63(5):E391–E392. https://doi.org/10.1503/cjs.011420

20. Felfeli T, Ximenes R, Naimark D, Hooper PL, Campbell RJ, El-Defrawy SR, Sander B (2021) The ophthalmic surgical backlog associated with the COVID-19 pandemic: a population-based and microsimulation modelling study. CMAJ Open 9(4):E1063–E1072. https://doi.org/10.9778/cmajo.20210145

21. Meneghini RM (2021) Techniques and strategies to optimize efficiencies in the office and operating room: getting through the patient backlog and preserving hospital resources. J Arthroplasty 36(7_suppl):S49–S51. https://doi.org/10.1016/j.arth.2021.03.010

22. Ljungqvist O, Nelson G, Demartines N (2020) The post COVID-19 surgical backlog: now is the time to implement enhanced recovery after surgery (ERAS). World J Surg 44:3197–3198. https://doi.org/10.1007/s00268-020-05734-5

23. Magennic P, Begley A, Dhariwal DK, Smith A, Hutchison I (2022) Oral and Maxillofacial Surgery (OMFS) Consultant Workforce in the UK: consultant numbers resulting from recruitment issues, pension pressures, changing job-plans, and demographics when combined with the COVID backlog in elective surgery, requires urgent action. Br J Oral Maxillofac Surg 60(1):14–19. https://doi.org/10.1016/j.bjoms.2021.10.011

24. Billig JI, Sears ED (2020) The compounding access problem for surgical care: innovations in the post-COVID era. Ann Surg 272(2):e47–e48. https://doi.org/10.1097/SLA.0000000000004085

25. Ting DSJ, Ang M, Mehta JS, Ting DSW (2019) Artificial intelligence-assisted telemedicine platform for cataract screening and management: a potential model of care for global eye health. Br J Ophthalmol 103(11):1537–1538. https://doi.org/10.1136/bjophthalmol-2019-315025

26. Ting DSJ, Foo VH, Yang LWY, Sia JT, Ang M, Lin H, Chodosh J, Mehta J, Ting DSW (2019) Artificial intelligence for anterior segment diseases: emerging applications in ophthalmology. Br J Ophthalmol 105(2):158–168. https://doi.org/10.1136/bjophthalmol-2019-315651

27. Simon MJK (2021) Regan WD (2021) COVID-19 pandemic effects on orthopaedic surgeons in British Columbia. J Orthop Surg Res 16:161. https://doi.org/10.1186/s13018-021-02283-y

28. Bleustein C (2020) Improving the Elective Surgery Backlog Caused by the Pandemic. Physicians Practice. https://www.physicianspractice.com/view/improving-the-elective-surgery-backlog-caused-by-the-pandemic. Accessed 2 May 2021

29. Lin P, Naveed H, Eleftheriadou M, Purbrick R, Ghanavati MZ, Liu C (2021) Br J Ophthalmol 105:745–750. https://doi.org/10.1136/bjophthalmol-2020-316917

30. Tzeng C, Teshome M, Katz M, Weinberg J, Lai S, Antonoff M, Bird J, Shafer A, Davis J, Adelman D, Moon B, Reece G, Prabhu S, DeSnyder S, Skibber J, Mehran R, Schmeler K, Roland C, Cao H, Aloia T, Caudle A, Swisher S, Vauthey J (2020) Cancer surgery scheduling during and after the COVID-19 first wave. Ann Surg 272(2):e106–e111. https://doi.org/10.1097/SLA.0000000000004092

31. Gregory A, Grant M, Boyle E, Arora R, Williams J, Salenger R, Chatterjee S, Lobdell K, Jahangiri M, Engelman D (2020) Cardiac surgery-enhanced recovery programs modified for COVID-19: Key steps to preserve resources, manage caseload backlog, and improve patient outcomes. J Cardiothorac Vasc Anesth 34(12):3218–3224. https://doi.org/10.1053/j.jvca.2020.08.007

32. Martin G, Clarke J, Markar S, Carter A (2020) How should hospitals manage the backlog of patients awaiting surgery following the COVID-19 pandemic? A demand modelling simulation case study for carotid endarterectomy. medRxiv preprint. https://www.medrxiv.org/content/10.1101/2020.04.29.20085183v1. Accessed 4 Apr 2021

33. Valente R, Di Domenico S, Mascherini M, Santori G, Papadia F, Orengo G, Gratarola A, Cafiero F, De Cian F (2021) A new model to prioritize waiting lists for elective surgery under the COVID-19 pandemic pressure. Br J Surg 108(1):e12–e14. https://doi.org/10.1093/bjs/znaa028

34. Matava C, So J, Williams R, Kelley S, ORRACLE-Xtra Group (2022) A Canadian Weekend Elective Pediatric Surgery Program to reduce the COVID-19–related backlog: Operating Room Ramp-Up After COVID-19 Lockdown Ends—Extra Lists (ORRACLE-Xtra) implementation study. JMIR Perioper Med 5(1):e35584. https://doi.org/10.2196/35584

35. Heyman DP, Sobel MJ (1984) Stochastic Models in Operations Research. Over Publications Inc, Mineola

36. Otterlo M, Wiering M (2012) Reinforcement learning and Markov decision processes. In: Wiering M, Otterlo M (eds) Reinforcement Learning: State-of-the-Art. Springer

37. Kaelbling LP, Littman ML, Moore AW (1996) Reinforcement learning: a survey. J Artif Intell Res 4:237–285. https://doi.org/10.1613/jair.301

38. Yu C, Liu J, Nemati S (2021) Reinforcement learning in healthcare: a survey. ACM Comput Surv 55(1):1–36. https://doi.org/10.1145/3477600

39. Zhao Y, Zeng D, Socinski MA, Kosorok MR (2011) Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. Biometrics 67(4):1422–1433. https://doi.org/10.1111/j.1541-0420.2011.01572.x

40. Padmanabhan R, Meskin N, Haddad WM (2015) Closed-loop control of anesthesia and mean arterial pressure using reinforcement learning. Biomed Signal Process Control 22:54–64. https://doi.org/10.1016/j.bspc.2015.05.013

41. Nemati S, Ghassemi MM, Clifford GD (2016) Optimal medication dosing from suboptimal clinical examples: a deep reinforcement learning approach. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). https://doi.org/10.1109/EMBC.2016.7591355

42. Padmanabhan R, Meskin N, Haddad WM (2017) Reinforcement learning-based control of drug dosing for cancer chemotherapy treatment. Math Biosci 293:11–20. https://doi.org/10.1016/j.mbs.2017.08.004

43. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA (2018) The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 24:1716–1720. https://doi.org/10.1038/s41591-018-0213-5

44. Sun Q, Jankovic MV, Budzinski J, Moore B, Diem P, Stettler C, Mougiakakou SG (2019) A dual mode adaptive basal-bolus advisor based on reinforcement learning. IEEE J Biomed Health Inform 23(6):2633–2641. https://doi.org/10.1109/JBHI.2018.2887067

45. Yu C, Ren G, Dong Y (2020) Supervised-actor-critic reinforcement learning for intelligent mechanical ventilation and sedative dosing in intensive care units. BMC Med Inform Decis Mak 20:124. https://doi.org/10.1186/s12911-020-1120-5

46. Padmanabhan R, Meskin N, Khattab T, Shraim M, Al-Hitmi M (2021) Reinforcement learning-based decision support system for COVID-19. Biomed Signal Process Control 68:102676. https://doi.org/10.1016/j.bspc.2021.102676

47. Fakih SJ, Das TK (2006) LEAD: A methodology for learning efficient approaches to medical diagnosis. IEEE Trans Inf Technol Biomed 10(2):220–228. https://doi.org/10.1109/TITB.2005.855538

48. Liu D, Jiang T (2018) Deep reinforcement learning for surgical gesture segmentation and classification. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, 11073. Springer, Cham. https://doi.org/10.1007/978-3-030-00937-3_29

49. Ghesu FC, Georgescu B, Zheng Y, Grbic S, Maier A, Hornegger J, Comaniciu D (2017) Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans. IEEE Trans Pattern Anal Mach Intell 41(1):176–189. https://doi.org/10.1109/TPAMI.2017.2782687

50. Ghesu FC, Georgescu B, Grbic S, Maier A, Hornegger J, Comaniciu D (2018) Towards intelligent robust detection of anatomical structures in incomplete volumetric data. Med Image Anal 48:203–213. https://doi.org/10.1016/j.media.2018.06.007

51. Loftus TJ, Filiberto AC, Li Y, Balch J, Cook AC, Tighe PJ, Efron PA, Upchurch GR Jr, Rashidi P, Li X, Bihorac A (2020) Decision analysis and reinforcement learning in surgical decision-making. Surgery 168(2):253–266. https://doi.org/10.1016/j.surg.2020.04.049

52. Boger J, Hoey J, Poupart P, Boutilier C, Fernie G, Mihailidis A (2006) A planning system based on Markov decision processes to guide people with dementia through activities of daily living. IEEE Trans Inf Technol Biomed 10(2):323–33. https://doi.org/10.1109/TITB.2006.864480

53. Bauer R, Gharabaghi A (2015) Reinforcement learning for adaptive threshold control of restorative brain-computer interfaces: a Bayesian simulation. Front Neurosci 9:36. https://doi.org/10.3389/fnins.2015.00036

54. Li K, Rath M, Burdick JW (2018) Inverse reinforcement learning via function approximation for clinical motion analysis. In 2018 IEEE international conference on robotics and automation (ICRA). IEEE: 610–617

55. Sahba F, Tizhoosh HR, Salama MM Application of opposition-based reinforcement learning in image segmentation. In 2007 IEEE symposium on computational intelligence in image and signal processing, IEEE: 246–251

56. Sahba F, Tizhoosh HR, Salama MM A reinforcement learning framework for medical image segmentation. In 2006 IEEE international joint conference on neural network proceedings, IEEE: 511–517

57. Netto SM, Leite VR, Silva AC, de Paiva AC, de Almeida Neto A (2008) Application on reinforcement learning for diagnosis based on medical image. In: Weber C, Elshaw M, Mayer NM (eds) Reinforcement Learning. InTech, Croatia

58. Jagodnik KM, Thomas PS, van den Bogert AJ, Branicky MS, Kirsch RF (2017) Training an actor-critic reinforcement learning controller for arm movement using human-generated rewards. IEEE Trans Neural Syst Rehabil Eng 25(10):1892–1905. https://doi.org/10.1109/TNSRE.2017.2700395

59. Pilarski PM, Dawson MR, Degris T, Carey JP, Sutton RS Dynamic switching and real-time machine learning for improved human control of assistive biomedical robots (2012) In 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), IEEE: 296–302

60. Daskalaki E, Scarnato L, Diem P, Mougiakakou SG (2010) Preliminary results of a novel approach for glucose regulation using an actor-critic learning based controller. In UKACC International Conference on CONTROL: 241–245. https://doi.org/10.1049/ic.2010.0287

61. Huang Z, van der Aalst W, Lu X, Duan H (2011) Reinforcement learning based resource allocation in business process management. Data Knowl Eng 70(1):127–145. https://doi.org/10.1016/j.datak.2010.09.002

62. Schutz H, Kolisch, (2012) Approximate dynamic programming for capacity allocation in the service industry. Eur J Oper Res 218(1):239–250. https://doi.org/10.1016/j.ejor.2011.09.007

63. Gomes TSMT (2017) Reinforcement learning for primary care appointment scheduling. Faculdade de Engenharia da Universidade do Porto

64. Lee S, Lee YH (2020) Improving emergency department efficiency by patient scheduling using deep reinforcement learning. Healthcare 8(2):77. https://doi.org/10.3390/healthcare8020077

65. Otterlo M, Wiering M (2012) Reinforcement learning and Markov decision processes. In Wiering M, Otterlo M (eds) Reinforcement Learning State-of-the-Art. Springer. https://doi.org/10.1007/978-3-642-27645-3

66. Heyman DP, Sobel MJ (2004) Stochastic Models in Operations Research. Dover Publications Inc, Mineola

67. Liu B, Xie Q, Modiano E (2019) Reinforcement learning for optimal control of queueing systems. 2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 24–27 September 2019, Monticello, IL, USA. IEEE. 663–670. https://doi.org/10.1109/ALLERTON.2019.8919665

68. Foss S, Shneer S, Tyurlikov A (2012) Stability of a Markov-modulated Markov chain, with application to a wireless network governed by two protocols. Stoch Syst 2(1):208–231. https://doi.org/10.1287/11-SSY030