Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

OPTIMIZING TASK SCHEDULING IN PRIMARY HEALTHCARE: A REINFORCEMENT LEARNING APPROACH WITH AGENT-BASED SIMULATION

Cristián Cárdenas[1], Gabriel Bustamante[3], Hernan Pinochet[3], Veronica Gil-Costa[2], Luis Veas-Castillo[1], and Mauricio Marin[3]

[1]Dept. of Informatics, Universidad Austral de Chile, Valdivia, CHILE
[2] Universidad Nacional de San Luis, San Luis, ARGENTINA
[3] Universidad de Santiago de Chile, Santiago, CHILE

ABSTRACT

The integration of Agent-Based Simulation (ABS) and Reinforcement Learning (RL) has emerged as a promising and effective approach for supporting decision-making in medical and hospital settings. This study proposes a novel framework that combines an Agent-Based Simulation with a Double Deep Q-Network (DDQN) Reinforcement Learning model to optimize task scheduling of healthcare professionals responsible for elderly patient care. Simulations were conducted over a 365-day period involving 250 patients, each managed by a healthcare coordinator who schedules appointments. Patients autonomously decide whether to attend appointments and adhere to medical recommendations. Results show the effectiveness of the RL model in minimizing health risks, with 84.8% of patients maintaining or improving their initial health risk levels, while only 15.2% experienced an increase.

1 INTRODUCTION

The healthcare systems worldwide face increasing challenges in providing efficient and high-quality care for the elderly population, particularly in primary healthcare settings (Jones and Dolsten 2024). The growing number of elderly patients with chronic conditions, such as hypertension, diabetes, and kidney disease, demands optimized patient management strategies to reduce clinical risks and avoid unnecessary hospitalizations (Beil et al. 2021). Traditional patient management approaches, which rely on manual task assignments and heuristic decision-making by healthcare professionals, often lead to inefficiencies, increased workloads, and a higher likelihood of errors (Rodziewicz et al. 2025). Consequently, there is a critical need for intelligent decision-support systems that can enhance patient care management (Alves et al. 2024; Kumar et al. 2018).

One of the major difficulties in addressing this problem is the dynamic and complex nature of healthcare environments. The need for personalized care, the variability in patient conditions, and resource constraints make it challenging to design a universally effective system. Simulation-based approaches have been widely used to model and analyze healthcare workflows, but their effectiveness is limited when it comes to optimizing real-time decision-making processes (Ruiz et al. 2024; Kasaie et al. 2018; Pepino et al. 2015; Almagooshi 2015). On the other hand, Reinforcement Learning (RL) has emerged as a promising technique for enhancing decision-making by learning optimal policies through trial and error, enabling adaptive responses to real-time changes in patient status, task urgency, and resource availability (Jayaraman et al. 2024; Ali 2022).

Previous studies have explored the use of RL in healthcare applications, demonstrating its potential in optimizing treatment strategies, patient scheduling, and clinical decision-making (Yu et al. 2021). However, the integration of RL within agent-based simulations to directly improve the task scheduling processes for healthcare professionals in primary healthcare centers has not been widely investigated, particularly to reduce the health risk of elder patients (Abdellatif et al. 2023; Allen and Monks 2020), mainly due to the challenges

in ethical, regulatory, and practical implementation of AI in clinical settings. Most existing approaches focus on either theoretical models or specific medical interventions without addressing the operational challenges of managing healthcare workflows dynamically (Ali 2022). This leads to the research question: Can Reinforcement Learning (RL) integrated into Agent-based Simulations (ABS) effectively optimize task scheduling for healthcare professionals to reduce clinical risk in elderly patients within primary care settings?

In this paper, we propose an RL-based approach that integrates a Double Deep Q-Network (DDQN) model (Van Hasselt et al. 2016) with an agent-based simulation to optimize task recommendation for healthcare professionals managing elderly patients. The proposed system learns optimal task scheduling strategies by interacting with a simulated environment that mimics real-world healthcare scenarios. This allows the model to dynamically adapt to changing patient conditions, prioritize critical cases, and reduce the overall clinical risk of patients. By leveraging RL, we aim to enhance the efficiency of healthcare providers while improving patient outcomes. The experimental results show that our RL-based system significantly improves task scheduling and allows to minimize the health risk of elderly patients. The use of DDQN enables adaptive decision-making, leading to a reduction in missed appointments and improving the prioritization of high-risk patients.

The content of this paper is organized as follows. Section 2 presents related work. Section 3 describes our case study. Section 4 presents our proposed agent-based simulation model and how we integrate it with a DDQN model. Section 5 presents the experimental results, and Section 6 concludes.

2 RELATED WORK

Recent literature reflects increasing interest in the application of ABS and RL to enhance healthcare systems. These studies range from the development of personalized treatments and chronic disease management to the optimization of treatment regimens and overall healthcare processes.

From a simulation perspective, ABS has proven to be a powerful tool for analyzing complex interactions within healthcare and elder care systems. The study by (Li et al. 2016) examines the interactions among patients, caregivers, and the environment in the context of chronic diseases such as diabetes, obesity, and heart disease, emphasizing both the clinical implications and societal relevance of the findings. In the context of aging populations, research such as that by (Büsing et al. 2020) demonstrates the potential of ABS to model doctor–patient interactions in rural settings, as well as the barriers older adults face in adopting medical appointment scheduling technologies—factors that frequently place them at a disadvantage compared to the broader population. Another ABS study, presented by (Prédhumeau and Manley 2025), utilizes data from COVID-19 vaccination campaigns and hospital records to evaluate the impact of vaccination delays on ICU occupancy and mortality rates. Collectively, these studies highlight the versatility and practical value of ABS for addressing complex challenges in contemporary healthcare systems.

Conversely, RL has demonstrated considerable potential to enhance clinical decision-making and the management of complex treatments within healthcare systems. In this context, (Abdellatif et al. 2023) presents an exhaustive review of over 150 RL applications, highlighting techniques aimed at developing personalized treatments and improving chronic disease care. Similarly, in the specific case of (Shortreed et al. 2011), RL is employed to improve prevention and treatment strategies for chronic conditions, promoting data-driven approaches to personalized medicine. Meanwhile, (Yu et al. 2021) investigates various applications of RL in healthcare, focusing on three core areas: edge intelligence, smart core networks, and dynamic therapeutic regimens. The study analyzes how RL can optimize diverse processes, ranging from treatment administration to the coordination of complex healthcare systems. Taken together, these contributions illustrate RL's potential as a powerful tool for supporting complex decision-making in healthcare environments.

Evidence from specific applications has validated the use of real-world data and DRL in healthcare. (Liu et al. 2019) employs DQN to design therapeutic sequences aimed at the prevention and treatment of graft-versus-host disease (GVHD).(Gottesman et al. 2019) addresses key challenges, including observational

bias in medical records, through the application of RL in clinical decision-making. Within the domain of mental health, (Xue et al. 2022) applies RL techniques to develop personalized interventions for patients with mental disorders, adapting treatment plans based on individual responses. These studies underscore the versatility and applicability of RL in supporting a wide range of clinical and healthcare-related tasks.

There are studies that have validated the integration of ABS and RL. For instance, (Lazebnik 2023) employs simulated environments and applies DRL to address the problem of hospital resource allocation, demonstrating superior performance compared to human decision-making. However, the study emphasizes that more realistic simulations are required to ensure clinical applicability. Similarly, (Abdullah et al. 2023) introduces an innovative approach that combines DRL with blockchain technology for task scheduling in healthcare systems, utilizing simulation as a testing framework. This solution addresses critical challenges related to security, scheduling, and operational costs in cloud-based healthcare environments, while fulfilling privacy and efficiency requirements in distributed networks. These studies underscore the feasibility and potential of integrating both techniques to achieve improved outcomes in healthcare contexts.

Our proposal differs from the reviewed studies in its practical and specific approach. While prior studies have focused on the use of RL to optimize disease treatments, our work focuses on implementing a DDQN model in a simulator of primary healthcare centers focused on elderly patients with chronic diseases. This includes medical treatment scheduling and clinical risk classification.

3 CASE STUDY

Case management in healthcare has become a key tool for supporting patients with high clinical or social risk, offering personalized and effective care based on objective assessments customized to the context of each patient (Hudon et al. 2015). It is a continuous process that involves planning, monitoring, and coordinating care for people with complex or chronic needs (Hernández-Zambrano et al. 2019), the duration depending on the condition of the patient and can extend over weeks, months, or even years (Müller et al. 2024).

In this model, the case manager, whether a nurse, physician, or social worker, plays a key role in the coordination of services between different levels of care and between specialists (Katz and Flarey 2024), improving clinical outcomes and reducing costs in the management of chronic diseases (Klaehn and Jaschke 2022). Their role emphasizes patient participation and empathetic communication, fostering adherence to treatment, satisfaction with care, and mobilizes the necessary resources to support patient well-being (Katz and Flarey 2024; Carr 2007).

The main challenge is maintaining patient stability and prevent deterioration, as significantly improving their clinical and social conditions is often difficult. This requires a case management framework that ensures balanced resource allocation and timely interventions. Studies have identified specific characteristics of primary care case management that are associated with positive outcomes in patients who frequently use healthcare services (Hudon et al. 2019). However, the implementation of such programs faces barriers that must be understood to ensure their success (Hacker et al. 2020).

In this scenario, the integration of DRL agents becomes a key element. DRL systems optimize resource allocation by balancing attention between high-risk and low-risk patients, supporting informed decision-making without compromising system stability. Research shows promising results in reducing mortality and improving fairness in resource distribution (Li et al. 2024).

The integration of case management with artificial intelligence seeks not only to prevent health decline but also to maintain patient stability and help to improve resource distribution in high-demand healthcare settings.

4 SIMULATION MODEL

4.1 Simulation overview

An Agent-Based Simulation (ABS) model is implemented to represent the dynamics of a hospital where a total of M case managers supervise P patients. The simulation runs for D virtual days.

Each case manager is randomly assigned $N = P/M$ non-transferable patients. The simulation begins with each manager setting the initial state configuration for their assigned patients. Then, the manager must select one of their patients to allocate medical appointment hours and perform actions aimed at preventing the deterioration of the patient's health status; this process is repeated indefinitely. Patient selection is determined by a DRL agent, and it is assumed that all manager actions are executed effectively.

As an ABS model, managers interact with patients by assigning medical hours and providing the necessary recommendations to help maintain stable health conditions. Moreover, patients may or may not attend their assigned appointments, where the attendance probability for any given patient is denoted as $P_A \sim \mathscr{N}_{[a,b]}(\mu,\sigma)$, where $\mathscr{N}_{[a,b]}$ is the truncated normal distribution on the interval $[a,b]$ (Burkardt 2023), with mean $\mu = \frac{a+b}{2}$ and standard deviation $\sigma = \frac{|a-b|}{4}$. The actions of managers and patients are governed by mutually assigned events controlled by a central monitor synchronization program. The simulation incorporates availability constraints for both managers and patients, taking into account their working hours, rest periods, and scheduled tasks.

4.2 Risk Model

Each patient p is associated with two types of risk: clinical ($R_C = R_C(p,t)$) and social ($R_S = R_S(p,t)$), both derived from a continuous risk function $\hat{R} = \hat{R}(p,t)$, which models a risk level for a given patient p and time t. This function is computed based on two components: a *baseline risk* ($\hat{R}_0 = \hat{R}_0(p,t)$) and an *evolution score* ($S = S(p,t)$).

The *baseline risk* is modeled as $\hat{R}_0 \sim B(\alpha,\beta)$, where B denotes the Beta distribution (Gupta and Nadarajah 2004). Parameters α and β depend on the patient's age and the number of chronic diseases he/she has. As age increases, the probability of being assigned a high risk also increases. Similarly, a higher number of chronic diseases increments the likelihood of a high risk level.

The risk *evolution score* is determined by three factors: patient attendance or absence to scheduled appointments (S_A), manager attention or inattention (S_W), and random events that affect patient well-being (S_R). In this way, the total evolution score is defined as $S = S_A + S_W + S_R$. Patients at higher risk are more likely to receive negative scores, especially if they are unattended for extended periods.

The risk function is defined as $\hat{R} = \min(\max(0, \hat{R}_0 - \upsilon S, 1)$, where υ is a parameter that controls the influence of the score on the final risk. Furthermore, the following classification is established: $\hat{R} \leq \frac{1}{3}$ corresponds to low risk ($R_C|R_S = R_{LOW}$), $\frac{1}{3} < \hat{R} \leq \frac{2}{3}$ represents medium risk ($R_C|R_S = R_{MEDIUM}$), and $\hat{R} > \frac{2}{3}$ indicates high risk ($R_C|R_S = R_{HIGH}$).

The score associated with the attendance to scheduled appointments is modeled as $S_A = \delta N_A$, where $N_A = N_{\text{attendances}} - N_{\text{absences}}$ represents the adherence value to the clinical or social treatment and δ is the improvement factor for adherence to appointments at the health center.

The score related to lack of attention is defined as $S_W = f(W_{p,t}) + f_{\text{acc}}(t)$, where $f(W_{p,t}) \leq 0$ represents a penalty function based on waiting time, and $f_{\text{acc}}(t) \leq 0$ corresponds to an accumulation function that aggregates the penalty over time. Intuitively, $f(W_{p,t})$ penalizes the system when a patient waits longer than a certain threshold, and $f_{\text{acc}}(t)$ captures the long-term effects of prolonged inattention by accumulating a portion of the penalty over time, even if the patient eventually receives care.

The penalty function for inattention is modeled as $f(W_{p,t}) = -\xi(\hat{R})\min(0, W_{p,t} - W_{min}(\hat{R}))$, where $W_{p,t}$ is the patient's waiting time since their last appointment, $\xi(\hat{R})$ is a monotonically increasing function that amplifies the impact of waiting time according to the risk level, and $W_{\min}(\hat{R})$ is a waiting time threshold beyond which the penalty is activated. The accumulation function is defined recursively as: $f_{acc}(0) = 0$,

and $f_{\text{acc}}(t) = \kappa(f_{\text{acc}}(\tau_p) + \max(0,\ f(W_{p,t}) + \psi))$, where τ_p is the last time the patient p had attention in the past; $\psi > 0$ and $\kappa \in [0,1]$ are parameters that regulate the persistence of the penalty.

The random score associated with sudden events that increase risk is modeled as:

$$S_R \leftarrow \begin{cases} X \sim U(a,b) & \text{with probability} \quad \zeta \\ 0 & \text{with probability} \quad (1-\zeta) \end{cases}$$

where U denotes the uniform distribution and $[a,b]$ is the range of possible values. This score is updated at the end of each simulation day and captures unexpected events that can negatively affect the patient's well-being. The use of a probabilistic assignment allows the model to introduce stochastic fluctuations in risk, simulating real-life scenarios such as acute health deterioration or sudden social crises that are not directly related to care quality.

Four types of appointments can be scheduled, grouped into two dimensions: clinical (medical appointments and diagnostic tests) and social (social appointments and psychological sessions). Each type has a scheduling probability P_S, independent of the others. Appointments are only assigned to patients with medium or high levels of clinical or social risk. Attendance at these appointments reduces the probability the patient reaches high risk. However, all patients remain susceptible to spontaneous increases in clinical or social risk due to adverse events.

4.3 Integration of simulation and DRL

The DRL agent is integrated into a recommendation system that suggests the next patient to be attended based on the current state known by the simulated healthcare center. Case managers constantly consult this system and perform their tasks exactly as instructed. During the simulation, case managers do not spend their time calculating patient risk or deciding who should be attended in the next iteration.

The integration of the DRL agent with the simulator is achieved through an intermediary API, which enables real-time information exchange between both systems. This API provides the necessary mechanisms for the simulator to send the current state of the environment, receive a recommended action, and subsequently report the resulting transitions from that action.

The recommendation system service is implemented as a multithreaded program segmented into three main components: DRL Model Serving, responsible for generating actions based on the received state of the simulation; DRL Data Gathering, responsible for collecting environment transitions; and DRL Model Training, where the deep learning model parameters are updated based on accumulated experience.

Figure 1 illustrates the interaction between the simulator and the DRL service. Communication follows a request-response structure, in which the simulator reports the environment state, the DRL service determines the best action, and this action is executed by the simulator. Subsequently, the resulting transitions are sent for storage and later used to train the model. This modular architecture allows the simulation process to be decoupled from model training, facilitating the scalability and optimization of DRL.

To reduce the latency caused by packet transmission through the server and its impact on exhaustive testing, a shared memory scheme with semaphores has been implemented. This approach minimizes communication latency between the simulator and the recommendation service, and is adopted as a first solution exclusively for model calibration.

4.4 Deep Reinforcement Learning Agent

The problem is modeled as a Markov Decision Process (MDP) (Li 2023), since the simulator provides full observability of patient states and does not rely on patient features involving uncertainty. In addition, the use of model-free DRL techniques (Li 2023) is proposed, allowing training without knowledge of the transition probabilities. This implies that only the set of *states*, *actions*, and *reward* function must be defined.

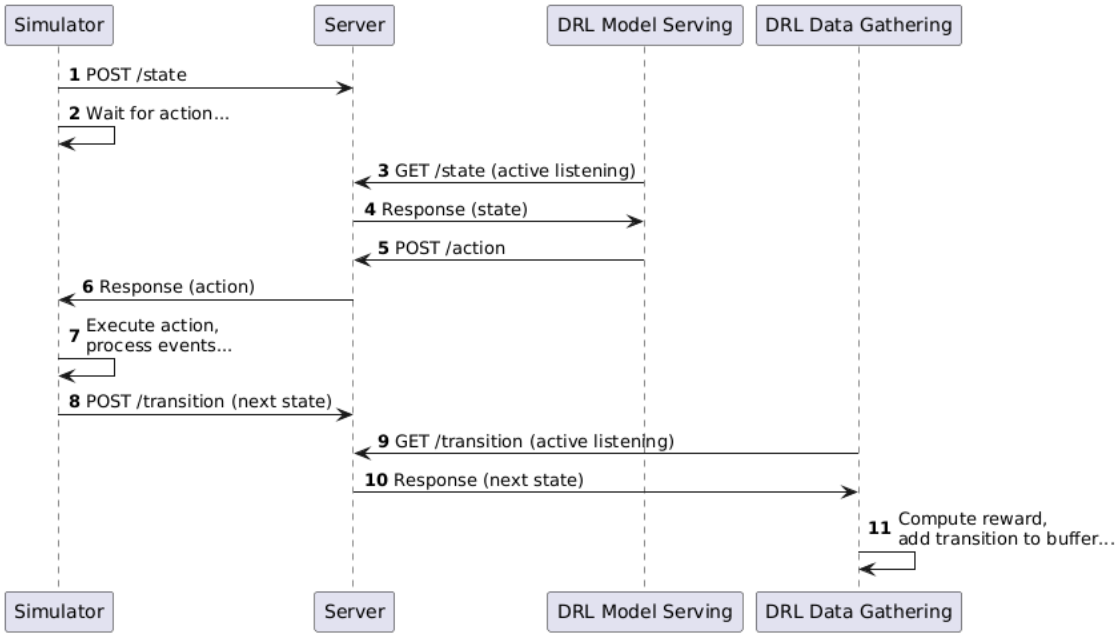

Figure 1: Interaction between the DRL Service and the Simulator

Since every manager is modeled independently from others, and the model is fed with experience coming from all managers, we use the tuple m, t to describe that a variable corresponds to manager m and is instantiated at time t. Thus, the *state* $s_{m,t} \in \mathbb{R}^{N \times c}$ corresponds to the concatenation between all N patients' feature vectors of size c from manager m at time t. The next state for manager m, given a state $s_{m,t}$ is denoted $s_{m,t+1}$.

The *action* $a_{m,t} \in [0,N]$ represents the index of the patient to be attended at that moment for manager m. This action is subsequently mapped to the id of the patient within the corresponding case manager's group, i.e. p can be obtained as a function of m and $a_{m,t}$.

The *reward* $r_{m,t}$ is defined as

$$r_{m,t} = R_{p,t} + \lambda W_{p,t} - \mu \bar{R}_{p,t+1} \tag{1}$$

where $W_{p,t} < W_{max}$ is the waiting time of the selected patient since their last appointment, $R_{p,t}$ is a function that combines clinical and social risks, calculated as

$$R_{p,t} = R_C + R_S + \eta |R_C - R_S| \tag{2}$$

and $\bar{R}_{p,t+1}$ is the average combined risk of all N patients in the next state. W_{max}, λ, μ, and η are hyperparameters that control the influence of each term on the reward.

In equation (1), to prevent the DRL agent from obtaining excessive rewards by indefinitely causing patient attention to delay, the waiting time $W_{p,t}$ is truncated by a maximum value W_{max}. This design encourages the agent to prioritize timely interventions while balancing global risk in the system. In equation 2, the term $\eta |R_C - R_S|$ is added to emphasize cases where there is a significant imbalance between clinical and social risks. This design results in a risk function that is more sensitive to extremes, assigning higher combined risk when either R_C or R_S approaches 1 even while the other remains low.

4.5 DDQN Approach

For the decision-making process of the DRL agent, a DDQN is employed, an enhanced variant of DQN that mitigates the overestimation of action values (Li 2023). In this context, the action values are directly

correlated with the urgency of each patient, enabling the agent to prioritize cases requiring more immediate attention. The adopted DDQN variant includes the use of Target Networks, which reduce the variance in action value estimation, and a Replay Buffer, which breaks the temporal correlation between successive samples (Li 2023).

As an exploration strategy, an ε-greedy scheme is adopted, in which the agent selects a random action with probability ε and an optimal action with probability $1-\varepsilon$. The exploration rate ε decays over time according to the function $\varepsilon(t) = \varepsilon_0 \varepsilon_{decay}^t$, where ε_0 is the initial exploration probability and ε_{decay} is a decay factor that governs the transition from exploration to exploitation of the learned policy. Both values are tunable hyperparameters.

To prevent the same patient from being selected multiple times on the same day, the action masking technique is applied (Huang and Ontañón 2022). This technique involves applying a restriction mask to the output neural network by assigning infinitely negative values to the action values corresponding to already selected patients. This ensures that a greedy policy with respect to the action values will never select restricted actions. The mask is applied both to the value-based selection and to randomly chosen actions within the ε-greedy scheme.

The neural network architecture used in the DDQN is based on convolutional neural networks (CNNs), as illustrated in Figure 2. The input state is represented by the matrix $s_{m,t} \in \mathbb{R}^{N \times c}$, over which k filters of size $(1 \times c)$ are applied to capture relevant information at the individual patient level. An activation function σ is then employed to introduce non-linearity into the representation. Finally, a second convolutional layer processes the output to estimate the action values before applying the action mask.

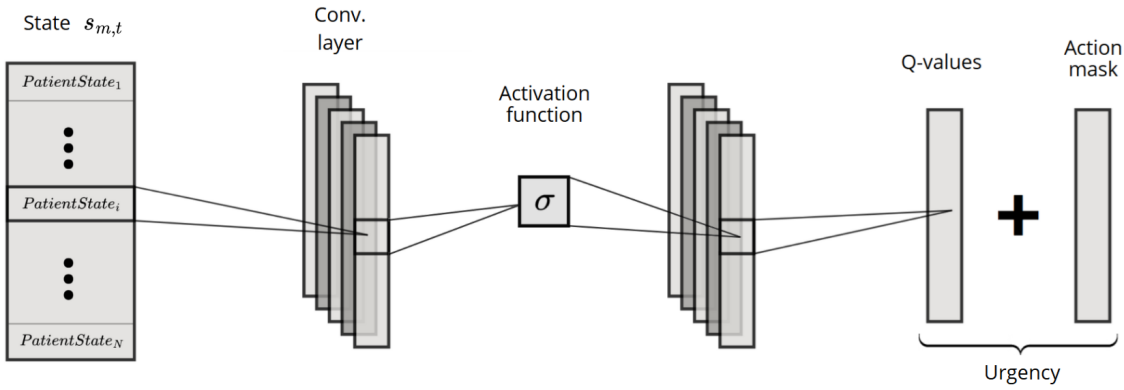

Figure 2: DDQN model architcture

5 EXPERIMENTS

5.1 Environment setup

The experiments are carried out with a total of $P = 250$ patients and $M = 5$ case managers over a continuous simulation period of $D = 365$ days. The probability that patients attend their scheduled appointments is modeled as $P_A \sim \mathcal{N}[a,b](\mu,\sigma)$, with $a = 0.46$ and $b = 0.62$ (μ and σ defined in Section 4.1). These parameters should be adjusted for each healthcare center based on historical patient attendance data. The simulation parameters used in Section 4.2 are: $\upsilon = 0.04$, $\delta = 1.0$, $\kappa = 0.9$, $\psi = 0.5$, $\zeta = 0.005$, $P_S = 0.6$. The distribution $\mathscr{U}(a,b)$ for S_R is defined in the interval $[a = -10, b = -5]$. For simplicity, the functions $\xi(x)$ and $W_{min}(x)$ are modeled using the categorical risk representation as follows: $\xi(R_{LOW}) = 0.025$, $\xi(R_{MEDIUM}) = 0.09$, $\xi(R_{HIGH}) = 0.025$ and $W_{min}(R_{LOW}) = 10$, $W_{min}(R_{MEDIUM}) = 7$, $W_{min}(R_{HIGH}) = 5$. The parameters used in Section 4.4 are $\lambda = 0.1$, $\mu = 1.0$, $\eta = 0.25$, and $W_{max} = 720$.

The features used to represent the state of each patient are: age (categorical variable with 3 classes), attended clinical hours (can take negative values), attended social hours, waiting time since the last

appointment, and number of chronic diseases. All numerical variables are normalized considering their respective maximum values (and minimums for those that may be negative). Based on this, the state size for each patient is $c = 7$ for the experiments in this work.

In the following subsections, the clinical risk (R_C) and the social risk (R_S) of the patients are represented numerically to facilitate visualization. Each risk category is assigned a specific value: $R_{LOW} = 10$, $R_{MEDIUM} = 20$, and $R_{HIGH} = 30$. In addition, all models are compared using the same random seed, which is generated at random before the beginning of the experiments. This process is repeated five times, averaging the results obtained for the patient's improvement rate, risk evolution, waiting time evolution, and average waiting time per average risk.

5.2 Results

5.2.1 Analysis of the behavior of the proposed approach

The results presented for the policy learned by the DRL agent (Figure 3) indicate that patients with higher risk are treated more frequently than those with lower risk. Although the frequency of care is not fixed for a given risk level, the reward function is shown to effectively associate risk with urgency of care, resulting in a policy capable of determining urgency without requiring risk as an explicit input.

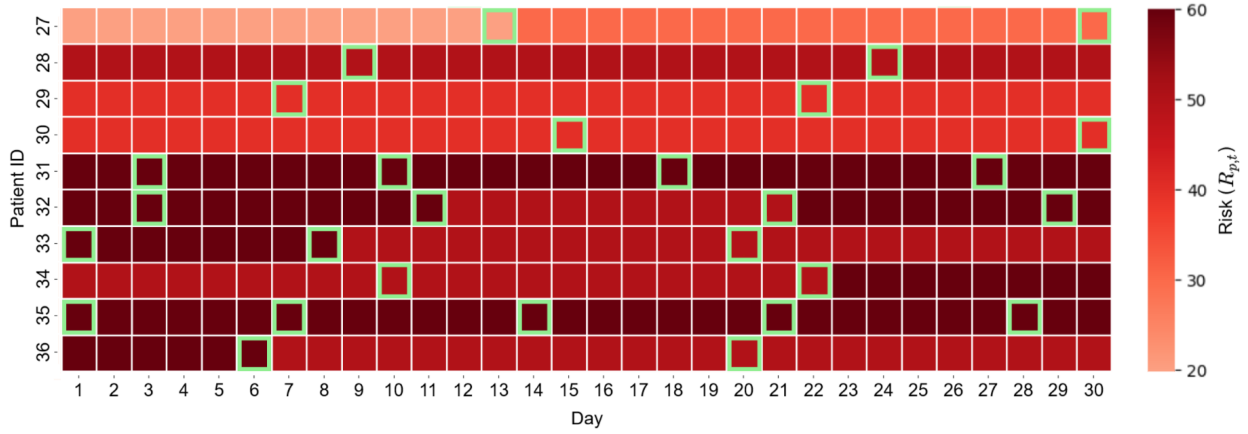

Figure 3: Visualization the patients health risk levels. The light green boxes indicate that the DRL agent selects a patient on the day specified on the x-axis. Upon selecting a given patient, their perceived risk will be updated in the following day.

Figure 4 illustrates the number of scheduled hours per patient (represented by their average combined risk $R_{p,t}$) over a simulated episode of 365 days, using only the policy learned by the DDQN agent. The figure shows an increase in the number of hours assigned for patients with higher risk levels, particularly those in the extreme category (high clinical and social risk), while healthier patients receive a lower number of hours. A low number of assigned hours does not necessarily imply that patients are not being constantly monitored by the case manager.

5.2.2 Evaluation of the proposed ABS-DRL approach

The results obtained from the simulation using the proposed DRL agent are compared against two baseline models. One baseline consists of randomly selecting patients, while the other aims to attend patients uniformly, which corresponds to the simulator's preliminary strategy. The latter method is equivalent to selecting the patient with the longest waiting time, making it a greedy algorithm with respect to $W_{p,t}$.

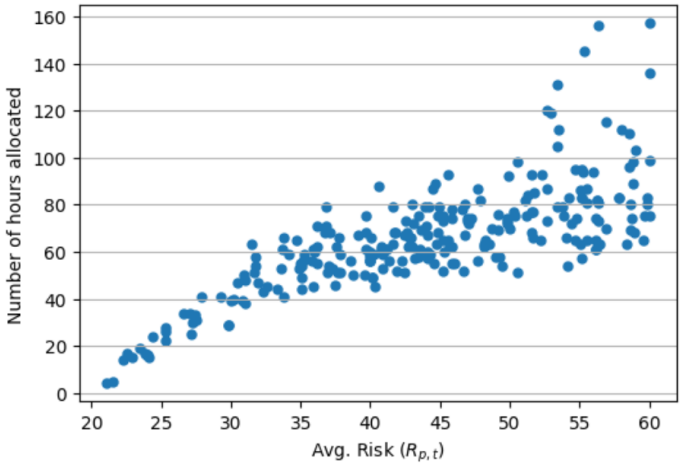

Figure 4: Scheduling of hours according to average risk for DDQN model

Table 1 summarizes the evolution of patients clinical and social risk levels ($R_{p,t}$), computed as the difference between their initial risk $R_{p,0}$ and final risk $R_{p,T}$ at the end of the simulation period. Values are averaged over five independent simulation runs. The DRL model clearly outperforms both baselines, with 50% of patients showing risk reduction, compared to 43.6% and 35.2% in the greedy and random policies, respectively. Additionally, the DRL model results in the lowest proportion of patients experiencing risk increases (15.2%), whereas the random baseline yields the highest (24.0%).

These findings suggest that the DRL agent is able to learn a policy that allocates resources more effectively, resulting in improved overall outcomes. The observed deltas, especially in the "Risk Decreases" and "Risk Increases" categories, highlight the added value of a learning-based strategy over static or heuristic-driven alternatives, enhancing both clinical and operational performance within the simulated environment.

Table 1: Average patient risk evolution ($R_{p,t}$) across 5 runs, comparing the proposed DRL model against baseline approaches.

Model	**Risk Decreases**	**Risk Unchanged**	**Risk Increases**
Proposed Model	125 (50%)	87 (34.8%)	38 (15.2%)
Greedy Baseline	109 (43.6%) [-6.4pp]	93 (37.2%) [+2.4pp]	48 (19.2%) [+4.0pp]
Random Baseline	88 (35.2%) [-14.8pp]	102 (40.8%) [+6.0pp]	60 (24.0%) [+8.8pp]

Figure 5 compares the proposed model with the two baselines based on three evaluation criteria. Plot (I) shows the daily evolution of the average combined risk ($R_{p,t}$) per patient. Plot (II) shows the average waiting time in days for each patient with respect to their risk level ($R_{p,t}$), where the behavior observed in the policy shown in Figure 3 can be appreciated. Finally, plot (III) illustrates the average time patients have been waiting since their last appointment. Patients who have been waiting for extended periods significantly affect the arithmetic mean.

5.2.3 Runtime Analysis

Figure 6 shows the simulation runtime as a function of different parameter combinations (N, D, and P). The plot on the left displays how runtime evolves with the total number of patients (P), keeping the simulation period fixed ($D = 365$) and varying the case manager workload (N). A nonlinear increase in runtime is observed as P grows, with significantly higher runtimes when the workload per manager is lower (i.e., smaller N). The plot on the right illustrates the effect of increasing the simulation duration (D), with a fixed number of patients ($P = 300$) and again varying N. In this case, runtime grows approximately linearly with

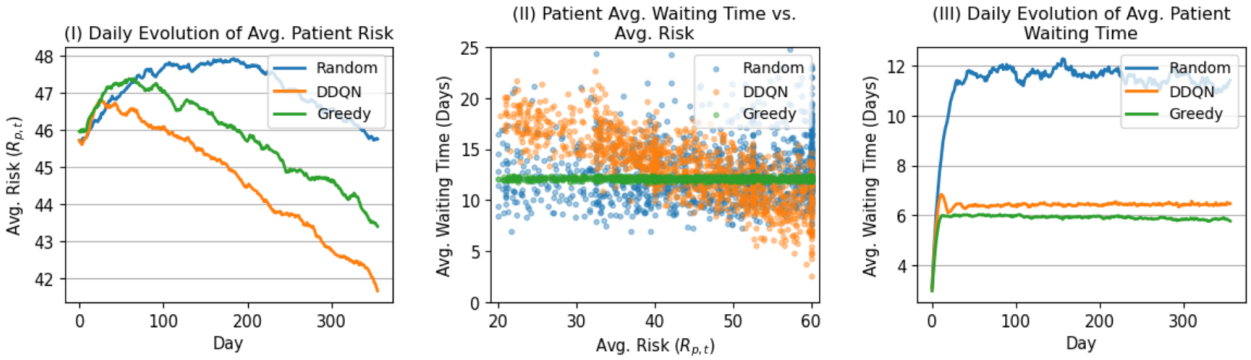

Figure 5: Comparison of the Model with Different Baselines.

D, and it is again evident that a lower workload per manager (smaller N) results in longer computation times.

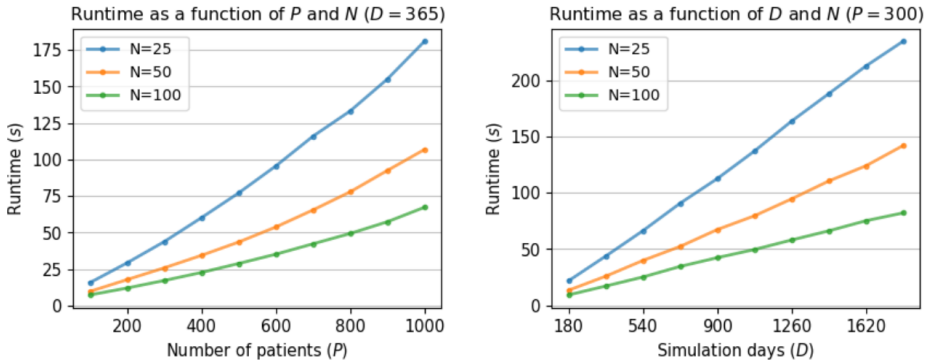

Figure 6: Simulation runtime using different values for the number of patients (P), number of days (D), and case manager workload (number of patients per manager, N).

6 CONCLUSIONS

This study shows the effectiveness of combining Agent-Based Simulation with Reinforcement Learning to support decision-making in primary healthcare systems. By modeling the appointment management process handled by case managers, the approach provided valuable insights into the dynamics and complexities of patient-caregiver interactions. The integration of a DDQN model allowed the simulation to optimize the scheduling of healthcare tasks while adapting to patient behavior, such as appointment attendance and adherence to medical advice.

The results suggest that this combination ABS-DDQN not only reduces patients health risks over time but also supports the efficient use of limited human resources in healthcare settings. The model proved capable of capturing the emergent behaviors of agents in a complex environment, making it a promising tool for evaluating and designing health interventions in realistic, data-driven scenarios.

As future work we plan to incorporate more granular patient data, such as comorbidities or behavioral profiles, could improve the realism and predictive power of the model. This would allow the system to better personalize the decision-making process for each patient and increase the model's capacity to simulate heterogeneous populations. Additionally, we plan to explore the integration of real-time data sources through database of health records, enabling the development of adaptive and dynamic simulations that reflect changes in patient status or resource availability.

7 ACKNOWLEDGMENTS

This work has been partially funded by the Agencia Nacional de Investigación y Desarrollo de Chile (ANID) under grant Basal Centre CeBiB code AFB240001.

REFERENCES

Abdellatif, A. A., N. Mhaisen, A. Mohamed, A. Erbad, and M. Guizani. 2023. "Reinforcement Learning for Intelligent Healthcare Systems: A Review of Challenges, Applications, and Open Research Issues". *IEEE Internet of Things Journal* 10(24):21982–22007.

Abdullah, L., M. M. Abed, N. Jan, M. Radek, T. Prayag, and K. Neeraj. 2023. "DRLBTS: Deep Reinforcement Learning-Aware Blockchain-Based Healthcare System". *Scientific Reports* 13(1):4124.

Ali, H. 2022. "Reinforcement Learning in Healthcare: Optimizing Treatment Strategies, Dynamic Resource Allocation, and Adaptive Clinical Decision-Making". *International Journal of Computer Applications Technology and Research* 11(3):88–104.

Allen, M., and T. Monks. 2020. "Integrating deep reinforcement learning networks with health system simulations". *arXiv preprint arXiv:2008.07434*.

Almagooshi, S. 2015. "Simulation Modelling in Healthcare: Challenges and Trends". *Procedia Manufacturing* 3:301–307.

Alves, M., J. Seringa, T. Silvestre, and T. Magalhães. 2024. "Use of artificial intelligence tools in supporting decision-making in hospital management". *BMC Health Services Research* 24(1):1–13.

Beil, M., H. Flaatten, B. Guidet, S. Sviri, C. Jung, D. de Lange, *et al*. 2021. "The management of multi-morbidity in elderly patients: Ready yet for precision medicine in intensive care?". *Critical Care* 25:1–7.

Burkardt, J. 2023. "The Truncated Normal Distribution". Technical report, Department of Scientific Computing, Florida State University, Tallahassee, Florida.

Büsing, C., S. Schmitz, M. Anapolska, S. Theis, M. Wille, C. Brandl, *et al*. 2020. "Agent-Based Simulation of Medical Care Processes in Rural Areas with the Aid of Current Data on ICT Usage Readiness Among Elderly Patients". In *Human Aspects of IT for the Aged Population. Healthy and Active Aging: 6th International Conference, ITAP 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22*, 3–12. Springer.

Carr, D. D. 2007. "Improving Transitions of Care: Bedside Report". *Nursing Management* 38(3):20–24.

Gottesman, O., F. Johansson, M. Komorowski, A. Faisal, D. Sontag, F. Doshi-Velez *et al*. 2019. "Guidelines for reinforcement learning in healthcare". *Nature Medicine* 25:16–18.

Gupta, A. K., and S. Nadarajah. 2004. *Handbook of Beta Distribution and Its Applications*. Boca Raton: CRC Press.

Hacker, M., I. Vedel, X. Q. Yang, E. Margo-Dermer, and C. Hudon. 2020. ""Understanding Barriers to and Facilitators of Case Management in Primary Care: A Systematic Review and Thematic Synthesis"". *Annals of Family Medicine* 18(4):355–363.

Hernández-Zambrano, S. M., I. Vargas, and M. L. Vázquez. 2019. ""Effectiveness of a Case Management Model for the Comprehensive Provision of Health Services to Multi-Pathological People"". *Journal of Advanced Nursing* 75(3):601–611.

Huang, S., and S. Ontañón. 2022. "A Closer Look at Invalid Action Masking in Policy Gradient Algorithms". *The International FLAIRS Conference Proceedings* 35.

Hudon, C., M.-C. Chouinard, F. Diadiou, M. Lambert, and D. Bouliane. 2015. ""Case Management in Primary Care for Frequent Users of Health Care Services With Chronic Diseases: A Qualitative Study of Patient and Family Experience"". *Annals of Family Medicine* 13(6):523–528.

Hudon, C., M.-C. Chouinard, P. Pluye, R. El Sherif, P. L. Bush, B. Rihoux, *et al*. 2019. ""Characteristics of Case Management in Primary Care Associated With Positive Outcomes for Frequent Users of Health Care: A Systematic Review"". *Annals of Family Medicine* 17(5):448–458.

Jayaraman, P., J. Desman, M. Sabounchi, G. N. Nadkarni, and A. Sakhuja. 2024. "A Primer on Reinforcement Learning in Medicine for Clinicians". *npj Digital Medicine* 7(1):337.

Jones, C. H., and M. Dolsten. 2024. "Healthcare on the brink: navigating the challenges of an aging society in the United States". *NPJ Aging* 10(1):16.

Kasaie, P., W. D. Kelton, R. M. Ancona, M. J. Ward, C. M. Froehle, and M. S. Lyons. 2018. "Lessons Learned From the Development and Parameterization of a Computer Simulation Model to Evaluate Task Modification for Health Care Providers". *Academic Emergency Medicine* 25(2):238–249.

Katz, J. M., and D. Flarey. 2024. "Understanding the Role of a Case Manager: Key Responsibilities". Technical report, American Institute of Health Care Professionals. accessed 28th March.

Klaehn, A.-K., and J. Jaschke. 2022. ""Cost-effectiveness of Case Management: A Systematic Review"". *The American Journal of Managed Care* 28(7):e218–e225.

Kumar, R., R. Tabassum, and M. Sabri. 2018. "A Study on Role of Intelligent Decision Support Systems in Healthcare". *SSRN Electronic Journal* 6(1):541–546.

Lazebnik, T. 2023. "Data-driven hospitals staff and resources allocation using agent-based simulation and deep reinforcement learning". *Engineering Applications of Artificial Intelligence* 126:106783.

Li, S. E. 2023. *Reinforcement Learning for Sequential Decision and Optimal Control*. Singapore: Springer.

Li, Y., M. A. Lawley, D. S. Siscovick, D. Zhang, and J. A. Pagán. 2016. "Agent-based modeling of chronic diseases: a narrative review and future research directions". *Preventing chronic disease* 13:E69.

Li, Y., C. Mao, K. Huang, H. Wang, Z. Yu, M. Wang *et al*. 2024. "Deep Reinforcement Learning for Efficient and Fair Allocation of Health Care Resources". *arXiv preprint arXiv:2309.08560*.

Liu, N., Y. Liu, B. Logan, Z. Xu, J. Tang, and Y. Wang. 2019. "Learning the Dynamic Treatment Regimes from Medical Registry Data through Deep Q-network". *Scientific reports* 9(1):1495.

Müller, A., F. Hebben, K. Dillen, V. Dunkl, Y. Goereci, R. Voltz, *et al*. 2024. ""So at least now I know how to deal with things myself, what I can do if it gets really bad again"—experiences with a long-term cross-sectoral advocacy care and case management for severe multiple sclerosis: a qualitative study". *BMC Health Services Research* 24(1):245.

Pepino, A., A. Torri, O. Tamburis *et al*. 2015. "Implementing simulation–based approaches for healthcare workflow analysis The Case of a Department of Laboratory Medicine in South Italy". In *proceedings of the 3rd International Conference on Advances in Computing, Communication and Information Technology-CCIT*.

Prédhumeau, M., and E. Manley. 2025. "Modéliser la mobilité des populations âgées pour guider un service de transport à la demande".

Rodziewicz, T. L., B. Houseman, S. Vaqar, and J. E. Hipskind. 2025. *Medical Error Reduction and Prevention*. Treasure Island: StatPearls Publishing, Treasure Island (FL).

Ruiz, M., E. Orta, and J. Sánchez. 2024. "A simulation-based approach for decision-support in healthcare processes". *Simulation Modelling Practice and Theory* 136:102983.

Shortreed, S. M., E. Laber, D. J. Lizotte, T. S. Stroup, J. Pineau, and S. A. Murphy. 2011. "Informing sequential clinical decision-making through reinforcement learning: an empirical study". *Machine learning* 84(1):109–136.

Van Hasselt, H., A. Guez, and D. Silver. 2016. "Deep reinforcement learning with double q-learning". In *Proceedings of the AAAI conference on artificial intelligence*, Volume 30.

Xue, Y., R. Farzan, and P. Brusilovsky. 2022. "Deep Reinforcement Learning for Personalized Adaptive Learning". *Journal of Educational and Behavioral Statistics* 48(2):220–243.

Yu, C., J. Liu, S. Nemati, and G. Yin. 2021. "Reinforcement learning in healthcare: A survey". *Association for Computing Machinery* 55(1):1 – 36.

AUTHOR BIOGRAPHIES

CRISTIÁN CÁRDENAS is a BSc student in Computer Science at Universidad Austral de Chile. His research interests include applied deep reinforcement learning in different areas such as simulation, NLP and robotics. His email address is cristian.cardenas01@alumnos.uach.cl.

GABRIEL BUSTAMANTE is Civil Engineering in Computer Science from Universidad de Santiago de Chile (USACH). His email address is gabriel.bustamante@usach.cl.

HERNAN PINOCHET is Civil Engineering in Computer Science from Universidad de Santiago de Chile (USACH). His email address is hernan.pinochet@usach.cl.

VERONICA GIL-COSTA received her Ph.D. in Computer Science, from Universidad Nacional de San Luis (UNSL), Argentina. She is a former researcher at Yahoo! Labs Santiago, hosted by the University of Chile. She is currently an Associate Professor at the University of San Luis and a researcher at the National Research Council (CONICET) of Argentina. Her email address is gvcosta@unsl.edu.ar.

LUIS VEAS-CASTILLO earned his Ph.D. in Engineering Sciences with a specialization in Computer Science from the University of Santiago, Chile. Currently, he is a full-time professor at the Austral University of Chile. His research areas include computational simulation, parallel and distributed systems, and web-based systems. His email address is luis.veasc@inf.uach.cl.

MAURICIO MARIN is a former researcher at Yahoo! Labs Santiago hosted by the University of Chile, and currently a full professor at University of Santiago, Chile. He holds a PhD in Computer Science from University of Oxford, UK, and a MSc from University of Chile. His research work is on parallel computing and distributed systems with applications in query processing and capacity planning for Web search engines. His email address is mauricio.marin@usach.cl.