ORIGINAL RESEARCH Open access

Hierarchical Reinforcement Learning Framework for Personalized Perioperative Antibiotic Prophylaxis Timing and Intraoperative Redosing

Ethan Wright[1], Chloe Bennett[1*], Jack Turner[2]

Abstract

Surgical site infections (SSIs) remain a significant source of postoperative morbidity despite established guidelines for perioperative antibiotic prophylaxis. Current protocols emphasize fixed preoperative timing and interval-based intraoperative redosing, yet fail to account for patient heterogeneity, pharmacokinetic variability, and uncertainty in procedure duration. This study proposes a hierarchical reinforcement learning (HRL) framework for personalized optimization of antibiotic prophylaxis across the perioperative timeline. The framework decomposes decision-making into two coordinated levels: a high-level policy that determines optimal preoperative antibiotic timing based on predicted procedure duration and patient-specific infection risk, and a low-level policy that adaptively manages intraoperative redosing using real-time updates on elapsed time, remaining duration, and cumulative drug exposure. Procedure duration is estimated using machine learning models that provide both point predictions and uncertainty intervals, enabling risk-sensitive decision-making. The problem is formalized as a Markov decision process with a reward structure balancing SSI prevention against antibiotic stewardship, incorporating penalties for unnecessary dosing and suboptimal timing. Off-policy evaluation using historical surgical data is proposed to assess performance relative to guideline-based and clinician-driven strategies. By integrating predictive modeling with multi-timescale decision optimization, the framework aims to reduce SSI incidence while minimizing antibiotic overuse. This approach highlights the potential of reinforcement learning to advance precision perioperative care and improve clinical outcomes through adaptive, data-driven prophylaxis strategies.

Keywords Surgical site infection, Hierarchical reinforcement learning, Antibiotic prophylaxis, Procedure duration prediction, Markov decision process, Perioperative decision support

*Correspondence:
Chloe Bennett
chloe.bennett@gmail.com
[1] Department of AI Healthcare Systems, University of Leeds, Leeds, United Kingdom
[2] Department of Clinical Intelligence Analytics, University of Sheffield, Sheffield, United Kingdom

Introduction

Surgical site infections affect an estimated 2% to 20% of patients undergoing surgical procedures, depending on wound classification, patient comorbidities, and procedure type, and represent one of the leading causes of postoperative readmission and healthcare expenditure [1, 2]. Perioperative antibiotic prophylaxis, when administered with appropriate timing relative to incision, reduces the incidence of SSIs by approximately 50% to 80% across a broad range of surgical specialties [3, 4]. The Surgical Care Improvement Project guidelines mandate that prophylactic antibiotics be administered within 60 to 120 minutes before surgical incision to achieve optimal tissue concentrations at the time of bacterial inoculation, a recommendation grounded in pharmacokinetic principles and observational evidence [5, 6]. Despite widespread adherence to these timing guidelines, SSIs persist at concerning rates,

suggesting that the current standardized approach may be insufficient for certain patient populations and procedure types.

Intraoperative redosing of prophylactic antibiotics is required for surgical procedures that exceed one to two times the drug's elimination half-life to maintain serum and tissue concentrations above the minimum inhibitory concentration for common surgical pathogens [7, 8]. For cefazolin, the most commonly used prophylactic agent with an elimination half-life of approximately 1.8 hours, redosing is typically recommended at four-hour intervals; however, this fixed schedule does not account for patient-specific factors such as body mass index, renal function, or estimated blood loss that influence drug pharmacokinetics [9, 10]. The decision to redose requires knowledge of both the elapsed procedure time and the anticipated remaining duration, information that is inherently uncertain and dynamically changing throughout the surgical course [11]. In practice, adherence to redosing recommendations is substantially lower than to preoperative timing guidelines, creating a vulnerable period during prolonged procedures when tissue antibiotic concentrations may fall below protective thresholds.

Procedure duration prediction represents a critical input to any adaptive prophylaxis strategy, yet surgical scheduling systems currently provide only rough estimates that frequently deviate from actual operative times [12, 13]. Machine learning models incorporating patient characteristics, procedure complexity codes, surgeon-specific factors, and institutional case volume have demonstrated superior predictive accuracy compared to historical averages or surgeon estimates alone [14, 15]. Furthermore, the communication of uncertainty in these predictions—providing not only a point estimate but also a prediction interval—enables downstream decision-making algorithms to adopt more conservative or aggressive strategies depending on confidence in the duration forecast [16, 17]. Integrating duration prediction with pharmacological modeling creates an opportunity for proactive, rather than reactive, antibiotic redosing decisions.

This article proposes a hierarchical reinforcement learning framework for personalized perioperative antibiotic prophylaxis that optimizes both the preoperative timing decision and the sequence of intraoperative redosing actions [18, 19]. The framework is conceptualized as a two-level architecture in which a high-level manager policy selects the optimal antibiotic administration time before incision based on predicted procedure duration and patient infection risk, while a low-level worker policy determines whether, when, and at what dose to redose intraoperatively given the elapsed time and remaining procedure duration [20, 21]. By framing prophylaxis optimization as a sequential decision problem across multiple timescales, this approach aims to reduce SSI rates while simultaneously minimizing unnecessary antibiotic exposure through personalized, adaptive administration schedules.

Background

Surgical site infection prevention

Contemporary SSI prevention guidelines from the Centers for Disease Control and Prevention and the Healthcare Infection Control Practices Advisory Committee specify several core measures related to antimicrobial prophylaxis, including SCIP-Inf-1 mandating that prophylactic antibiotics be received within one hour before surgical incision, SCIP-Inf-2 requiring appropriate antibiotic selection based on procedure type and patient allergies, and SCIP-Inf-9 stipulating discontinuation of prophylactic antibiotics within 24 hours of surgery completion [1, 8]. These process measures have been widely adopted as quality metrics and have driven substantial improvements in guideline adherence across surgical services. However, the evidence underlying specific timing thresholds derives primarily from observational studies rather than randomized trials, and the optimal preoperative window may vary by antibiotic agent, patient pharmacokinetics, and procedure-specific contamination risk [6, 9]. The 2017 CDC guideline update acknowledged the limitations of the existing evidence base, particularly regarding the need for redosing during prolonged procedures and the role of patient-specific risk stratification in determining prophylaxis intensity.

Antibiotic pharmacokinetics

The pharmacological rationale for perioperative antibiotic prophylaxis centers on achieving and maintaining tissue concentrations exceeding the minimum inhibitory concentration for likely contaminating organisms throughout the period from incision to wound closure [3, 7]. Cefazolin, a first-generation cephalosporin with a beta-lactam ring, exhibits time-dependent bactericidal activity with an elimination half-life of approximately 1.5 to 2.2 hours in patients with normal renal function, necessitating redosing at approximately four hours for prolonged procedures [10,

11]. Vancomycin, employed for patients with known methicillin-resistant Staphylococcus aureus colonization or those at high risk, demonstrates a longer half-life of four to six hours but requires extended infusion times of sixty minutes to avoid infusion-related reactions, complicating preoperative timing decisions [5, 22]. The relationship between serum concentration and tissue penetration varies across antibiotic classes and is influenced by patient factors including obesity, which alters volume of distribution, and renal function, which determines elimination kinetics.

Procedure duration prediction

Accurate prediction of surgical case duration has emerged as a critical operational and clinical challenge, with implications extending from operating room scheduling efficiency to prophylaxis decision support [12, 14]. Traditional approaches relying on surgeon-provided estimates or historical averages for procedure-surgeon combinations demonstrate systematic bias and limited precision, particularly for long-tail cases that substantially exceed predicted durations [15, 17]. Machine learning models trained on electronic health record data have demonstrated improved accuracy by incorporating features such as surgical Current Procedural Terminology codes, patient age and body mass index, American Society of Anesthesiologists physical status classification, surgeon historical case volume, and scheduled procedure start time [16, 20]. A randomized clinical trial evaluating the effect of a predictive model on planned surgical duration accuracy demonstrated significant reductions in patient wait times and more efficient use of presurgical resources when predictions were communicated to operating room teams [18]. Modular artificial neural network architectures capable of continuous real-time prediction updating as procedures progress offer the potential for dynamic duration forecasts that can inform intraoperative decision-making [19].

Hierarchical reinforcement learning

Hierarchical reinforcement learning addresses the challenge of sequential decision-making across multiple temporal and spatial scales by decomposing complex tasks into hierarchies of subtasks, enabling more efficient exploration and credit assignment than flat reinforcement learning approaches [23, 24]. The options framework formalizes temporally extended actions that execute over multiple primitive time steps, allowing a high-level policy to select among options while low-level policies execute the selected option until termination [25, 26]. Feudal reinforcement learning architectures implement a manager-worker structure in which the manager sets goals or subgoals for the worker, which learns to achieve those goals through primitive actions, with the manager receiving reward based on the worker's performance [27, 28]. This decomposition is particularly well-suited to medical domains where decisions occur on distinct timescales, such as the hours-before-surgery preoperative antibiotic timing decision and the minutes-during-surgery redosing decision, each of which requires a different temporal abstraction for effective learning [21]. The transferability of reinforcement learning models across healthcare settings has been demonstrated in critical care applications, suggesting the feasibility of learning generalizable policies from institutional data [29].

Framework Overview

High-level architecture

The proposed framework consists of four integrated components arranged in a sequential processing pipeline: a patient feature extractor that compiles relevant clinical characteristics and surgical context, a procedure duration predictor that generates uncertainty-aware estimates of operative time, a hierarchical reinforcement learning module comprising high-level and low-level policies, and a prophylaxis schedule generator that outputs actionable administration recommendations [17, 20]. Patient factors—including age, body mass index, diabetes status, MRSA colonization history, immunosuppression status, and American Society of Anesthesiologists classification—are combined with surgery type encoded by Current Procedural Terminology code to form the initial state representation [1, 5]. The duration predictor processes these inputs and generates both a point estimate and a prediction interval, which are passed to the high-level policy along with patient infection risk features [12, 14]. The high-level policy selects a preoperative antibiotic timing action, after which the low-level policy assumes control during the intraoperative period, making redosing decisions based on elapsed time, remaining predicted duration, and cumulative antibiotic exposure [21, 24].

Figure 1 illustrates the proposed hierarchical reinforcement learning architecture for coordinating uncertainty-aware preoperative antibiotic timing with adaptive intraoperative redosing decisions.

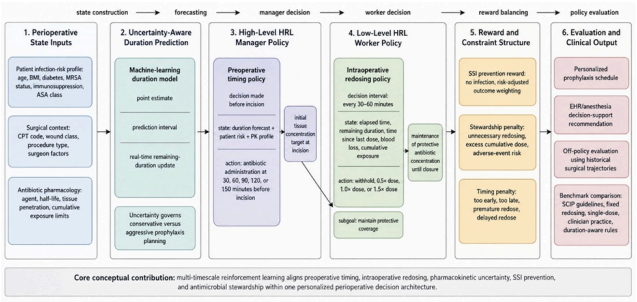

Figure 1. Hierarchical Reinforcement Learning Architecture for Personalized Perioperative Antibiotic Timing and Intraoperative Redosing

Core assumptions

This framework assumes the availability of historical surgical case data containing procedure durations, antibiotic administration times—both preoperative and intraoperative—patient demographics and comorbidities, and postoperative SSI outcomes as defined by National Healthcare Safety Network surveillance criteria [8, 9]. The data are assumed to capture sufficient variation in prophylaxis practices to enable learning of counterfactual policies through off-policy evaluation methods [25, 26]. A further assumption is that the causal relationship between tissue antibiotic concentration and SSI risk is mediated through pharmacokinetic principles that can be approximated by the time since last dose relative to the drug's elimination half-life [3, 7]. The framework also assumes that clinicians will comply with policy recommendations at a sufficiently high rate to generate informative training data, acknowledging that real-world adherence to decision support is imperfect and must be modeled explicitly [18, 22].

Design principles

Three design principles guide the development of this hierarchical reinforcement learning framework. First, multi-timescale optimization acknowledges that the preoperative timing decision and intraoperative redosing decisions operate on fundamentally different temporal horizons and should be learned by policies operating at appropriate levels of temporal abstraction [24, 27]. Second, duration prediction integration ensures that both policy levels have access to predicted procedure length and its associated uncertainty, enabling the high-level policy to make anticipatory timing decisions and the low-level policy to plan redosing sequences that account for remaining operative time [16, 19]. Third, personalized risk stratification recognizes that the optimal prophylaxis strategy for a patient with multiple SSI risk factors differs from that for a low-risk patient, and the reward function must reflect this heterogeneity by weighting outcomes according to baseline infection probability [2, 6].

Procedure Duration Prediction

Prediction model

The duration prediction model employs gradient-boosted tree algorithms or modular neural network architectures trained on historical surgical case data to estimate operative time from patient and procedural features [14, 16]. Input features include procedure Current Procedural Terminology code, patient age, body mass index, American Society of Anesthesiologists physical status classification, surgeon identifier and historical case volume, scheduled start time capturing circadian variation in operating room efficiency, and wound class as a proxy for procedural complexity [12, 15]. The model outputs a predicted duration in minutes along with an uncertainty interval derived from quantile regression or ensemble variance, providing downstream components with information about both the expected value and the confidence in that expectation [19, 20]. Continuous real-time updating of the prediction as the procedure progresses, using elapsed time and surgical milestone completion as additional features, enables dynamic revision of the duration forecast during the intraoperative period [17, 18].

Uncertainty-aware input to RL

The predicted duration and its associated uncertainty are communicated to the hierarchical reinforcement learning policies through the state representation, enabling risk-sensitive decision-making that adapts to prediction confidence [13, 21]. When the prediction interval is narrow—indicating high confidence in the duration estimate—the high-level policy can make a more aggressive preoperative timing decision, potentially administering antibiotics closer to the expected incision time to maximize tissue concentrations at the critical moment of bacterial exposure [3, 11]. Conversely, when the prediction interval is wide—reflecting substantial uncertainty about procedure length—the high-level policy may adopt a conservative approach, administering antibiotics earlier or selecting agents with longer half-lives to ensure coverage even if the procedure

extends well beyond the expected duration [1, 7]. The low-level policy similarly conditions its redosing decisions on the evolving uncertainty in remaining procedure time, becoming more likely to redose when the upper bound of the prediction interval exceeds the antibiotic's effective coverage period.

Hierarchical RL Architecture

High-level policy (Manager)

The high-level policy operates as a manager within a feudal reinforcement learning architecture, making a single decision at the time of surgical scheduling or immediately before the procedure regarding the optimal timing of preoperative antibiotic administration [21, 24]. The state space for the high-level policy includes the predicted procedure duration with its uncertainty interval, patient-specific infection risk features—such as MRSA colonization status, diabetes mellitus, immunosuppression, and obesity—and the selected antibiotic agent with its pharmacokinetic profile, particularly elimination half-life and time to peak tissue concentration [5, 10]. The action space consists of discrete preoperative timing options relative to scheduled incision: administration at 30, 60, 90, 120, or 150 minutes before the anticipated start time, with the understanding that tissue concentrations reach therapeutic levels approximately 30 minutes after infusion for most beta-lactam antibiotics [2, 6]. The high-level policy receives reward at the end of the surgical episode based on SSI outcome and antibiotic stewardship metrics, with credit assigned through the temporal abstraction mechanism to this preoperative decision.

Low-level policy (Worker)

The low-level policy functions as a worker that executes the intraoperative phase of antibiotic management, receiving a subgoal from the high-level policy to maintain serum antibiotic concentrations above the minimum inhibitory concentration for the duration of the procedure [24, 27]. At each intraoperative decision point—occurring every 30 to 60 minutes depending on institutional workflow and the specific antibiotic's pharmacokinetics—the low-level policy observes the current state, including time elapsed since the last antibiotic dose, cumulative antibiotic exposure, estimated remaining procedure duration from the continuously updating prediction model, and any patient physiological changes such as estimated blood loss exceeding 1500 milliliters that may alter pharmacokinetics [11, 14]. The action space includes the decision to withhold redosing or to administer a redose at 0.5 times, 1.0 times, or 1.5 times the standard weight-adjusted dose, enabling both dose reduction for patients approaching cumulative toxicity thresholds and dose escalation for those with demonstrated rapid clearance or ongoing hemorrhage [9, 23]. The low-level policy receives immediate rewards shaped to encourage maintenance of protective antibiotic concentrations while penalizing unnecessary redosing that contributes to antimicrobial resistance and adverse events [28, 29].

Temporal abstraction

Temporal abstraction is implemented through a feudal reinforcement learning structure in which the high-level policy selects a goal at the procedure outset and the low-level policy pursues that goal through a sequence of primitive actions spanning the intraoperative period [24, 25]. The high-level policy's decision—the preoperative timing action—is executed once and determines the initial antibiotic concentration at incision, setting the starting conditions for the low-level policy's subsequent redosing decisions [3, 7]. The low-level policy operates on a finer timescale, making decisions at regular intervals throughout the procedure, with the temporal abstraction mechanism ensuring that the high-level policy receives credit or blame for SSI outcomes based on the entire trajectory rather than any single intraoperative decision [26, 27]. This decomposition enables the high-level policy to learn long-horizon strategies for preoperative timing while the low-level policy specializes in the reactive, moment-to-moment management of antibiotic concentrations during the inherently uncertain course of surgery.

Table 1 decomposes the proposed framework into distinct decision-theoretic layers, clarifying how each component contributes to personalized prophylaxis optimization.

Table 1. Decision-Theoretic Decomposition of the Hierarchical Reinforcement Learning Framework

Framework layer	Clinical decision problem	Temporal scale	State inform used
Patient-state construction	Define individualized prophylaxis context	Before surgery	Age, BMI, dia MRSA sta immunosuppr ASA class, w

			class, proce type
Duration-prediction module	Estimate likely procedure length and remaining operative time	Scheduling to intraoperative period	CPT code, su history, proc complexi scheduled sta elapsed ti surgical miles
High-level manager policy	Select preoperative antibiotic timing	Single decision before incision	Predicted du prediction int patient SSI antibiotic ha time to pe concentra
Low-level worker policy	Manage intraoperative redosing	Repeated 30–60 minute decisions	Time since dose, elapse remaining pre duration, cum exposure, b loss, estim concentra
Reward structure	Balance infection prevention and stewardship	End-of-episode and shaped intermediate rewards	SSI outcor cumulative adequacy concentra coverage, ti deviation
Off-policy evaluation	Estimate policy value before deployment	Retrospective evaluation	Historica trajectorie behavior-p actions, outc dosing rec

MDP Formulation

State space

The Markov decision process state representation integrates patient-level, procedural, pharmacological, and temporal features into a unified vector that both the high-level and low-level policies observe at their respective decision points [1, 5]. Patient features include age, body mass index, diabetes status, MRSA colonization history, immunosuppression status, and American Society of Anesthesiologists physical status classification, each of which has been associated with differential SSI risk and altered antibiotic pharmacokinetics [2, 6]. Procedural features encompass the scheduled surgery type encoded by Current Procedural Terminology code, wound classification, and the predicted procedure duration with its associated uncertainty interval as generated by the machine learning prediction module [12, 14]. Pharmacological state variables capture the time elapsed since the last antibiotic dose, the cumulative antibiotic dose administered thus far, the estimated current serum concentration based on a one-compartment pharmacokinetic model, and the selected agent's elimination half-life [10, 11]. Temporal features indicating the current phase of care—preoperative versus intraoperative—and the elapsed procedure time enable the policies to condition their decisions on the progress of the surgical episode [17, 20].

Action space

The action space is structured hierarchically to reflect the distinct decision types made at different temporal scales, with the high-level policy selecting from a discrete set of preoperative timing options and the low-level policy choosing among intraoperative redosing actions [21, 24]. For the high-level preoperative timing decision, the action set consists of five discrete options corresponding to antibiotic administration at 30, 60, 90, 120, or 150 minutes before the scheduled incision time, with the choice of antibiotic agent assumed to follow institutional guidelines based on procedure type and patient allergy profile [1, 8]. The low-level intraoperative action space includes a discrete set of four actions: withhold redosing, administer a reduced redose at 0.5 times the standard weight-adjusted dose, administer a standard redose at 1.0 times the weight-adjusted dose, or administer an escalated redose at 1.5 times the standard dose to compensate for rapid clearance or substantial blood loss [3, 9]. This discrete action parameterization balances clinical interpretability with sufficient granularity to enable personalized dose adjustment, avoiding the complexity of continuous dose

selection while accommodating clinically meaningful dose modifications [23, 28].

Reward Design

SSI risk reduction reward

The primary reward incentivizes prevention of surgical site infections (SSIs), assigning positive reward for no SSI and penalties for SSI occurrence, weighted by baseline patient risk to avoid bias against high-risk cases [2, 6]. Baseline risk is estimated from preoperative models incorporating comorbidities, procedure type, and wound class, ensuring outcome significance is appropriately scaled [5, 22]. This risk-stratified design prioritizes high-risk patients, where optimized prophylaxis yields the greatest marginal benefit, supporting personalized decision-making [3, 7].

Antibiotic stewardship penalties

Stewardship penalties offset SSI-focused rewards by discouraging unnecessary antibiotic use that contributes to resistance, adverse events, and costs [28, 29]. Penalties are applied for redosing when drug levels remain adequate and additional coverage is unwarranted given remaining procedure time [9, 11]. Additional penalties scale with cumulative dose relative to recommended limits, reducing risks such as *Clostridioides difficile* infection and nephrotoxicity [8, 10]. Penalty magnitudes are calibrated to allow increased use when clinically justified for high-risk or prolonged cases [4, 23].

Timing penalty

Timing penalties enforce pharmacokinetic principles by discouraging administration outside optimal windows [1, 3]. Preoperative dosing too early (>120 minutes) or too late (<30 minutes) relative to incision incurs penalties due to reduced efficacy or insufficient tissue penetration [5–7, 22]. Intraoperatively, penalties apply to premature redosing (excess exposure) and delayed redosing (loss of protective coverage) [2, 11]. These constraints guide policies toward maintaining effective concentrations while avoiding unnecessary peaks.

Clinical Integration

Preoperative decision support

Integration of the hierarchical reinforcement learning framework into clinical workflow begins at the time of surgical scheduling, when the system generates an initial prophylaxis recommendation including the optimal antibiotic agent—if not already specified by institutional protocol—and the recommended preoperative administration timing based on the predicted procedure duration and patient risk profile [12, 18]. This recommendation is communicated to the surgical team and anesthesia providers through the electronic health record or operating room scheduling system, with the timing recommendation updated if the predicted procedure duration changes substantially due to scheduling modifications, surgeon reassignment, or updates to the patient's clinical status [17, 20]. The preoperative decision support interface displays not only the recommended timing but also the rationale in terms of predicted duration, antibiotic pharmacokinetics, and patient-specific infection risk, enabling clinicians to understand and evaluate the recommendation within their existing clinical knowledge framework rather than treating it as an opaque algorithmic output [14, 19]. As the time of surgery approaches and the scheduled start time becomes firmer, the preoperative timing recommendation becomes increasingly actionable, ultimately specifying the exact clock time at which the antibiotic infusion should be initiated to achieve the target interval before incision [1, 8].

Intraoperative alert system

During the intraoperative phase, the low-level policy's redosing recommendations are delivered through an alert system integrated with the anesthesia information management system, which captures real-time data on elapsed procedure time, administered medications, and patient physiological parameters [9, 11]. At each predefined decision interval—every 30 minutes for antibiotics with short half-lives or every 60 minutes for agents with longer half-lives—the system evaluates the current state, including the continuously updated remaining duration prediction, and generates a recommendation to either withhold redosing or administer a redose at a specified dose [14, 17]. The alert is designed to be minimally disruptive, appearing as a non-interruptive notification on the anesthesia workstation display rather than an audible alarm, with the acknowledgment and response of the anesthesia provider captured for subsequent policy evaluation and refinement [18, 22]. Critically, the system does not mandate compliance but rather functions as a clinical decision support tool, with the final decision resting with the anesthesia provider who may override the

recommendation based on clinical judgment regarding factors not captured in the state representation, such as ongoing hemodynamic instability or concern for anaphylaxis [5, 28].

Evaluation Strategy

Off-policy evaluation

Evaluation before prospective deployment employs off-policy methods that estimate the target hierarchical reinforcement learning policy's performance using historical data collected under standard care [25, 26]. Importance sampling reweights observed outcomes by the action probability ratio between target and behavior policies, while weighted importance sampling reduces variance for typical single-institution dataset sizes [24, 28]. The sequential nature of intraoperative redosing decisions requires per-decision or stepwise importance sampling to appropriately weight trajectories with multiple decision points [23, 25].

Performance metrics

Evaluation employs multi-dimensional metrics capturing both infection prevention and antibiotic stewardship outcomes [2, 7]. The primary clinical metric is estimated SSI rate reduction under the target policy, expressed as absolute and relative risk reduction [3, 6]. Stewardship metrics include guideline-appropriate redosing proportion, cumulative antibiotic dose, and rates of unnecessary and missed redoses [9, 10]. Additional process metrics assess preoperative timing adherence and intraoperative redosing interval compliance, while economic metrics estimate incremental costs incorporating antibiotic acquisition and SSI treatment expenditures [1, 4, 8, 11].

Baseline comparisons

The target policy is compared against clinically relevant baselines evaluated on the same historical dataset [25, 27]. The primary baseline follows Surgical Care Improvement Project guidelines with fixed-interval redosing, while a fixed-redosing strategy applies scheduled redoses without considering remaining duration [1, 3, 8, 9]. A single-dose-only strategy establishes a lower antibiotic exposure bound [6, 11]. A clinician-determined strategy uses documented decisions as the behavior policy, and a duration-aware rule-based strategy selects fixed prophylaxis schedules from predicted duration, representing the best non-learning alternative [14, 17, 18, 20].

Table 2 specifies the evaluation logic needed to determine whether the HRL target policy improves infection prevention, stewardship, uncertainty handling, and clinical deployability relative to existing prophylaxis strategies.

Table 2. Evaluation Logic for Comparing HRL-Based Prophylaxis Against Clinical Baselines

Evaluation dimension	HRL target policy expectation	Guideline-based baseline	Clini prac bas
Preoperative timing	Personalized timing selected from patient risk, antibiotic pharmacokinetics, and predicted procedure duration	Fixed administration window before incision	Variabl based workfl prov judg
Intraoperative redosing	Adaptive redosing based on elapsed time, remaining duration, exposure, and patient-specific factors	Fixed interval redosing based mainly on antibiotic half-life	Red deper prov aware docume a intraop work
SSI prevention	Risk-weighted reduction in postoperative SSI	Expected protection when guidelines are followed	Real- effecti un heterog adhe
Antibiotic stewardship	Minimized cumulative dose unless justified by risk or prolonged duration	Standard exposure may be excessive for short or low-risk cases	Exposu be inco acr provid ca
Robustness to duration	More conservative	Does not explicitly use	Deper informa

uncertainty	decisions when prediction intervals are wide; more precise decisions when confidence is high	predictive uncertainty	anticip case
Clinical deployability	Recommendation must be interpretable, overrideable, and compatible with EHR/anesthesia systems	Already embedded in institutional protocols	Alre embec routine
Retrospective safety assessment	Off-policy evaluation estimates policy value before implementation	Serves as primary reference standard	Serv obse behavi

Limitations

Technical limitations

Technical limitations center on duration prediction uncertainty and off-policy evaluation challenges in healthcare settings. Procedure duration prediction remains uncertain for long-tail cases deviating from historical patterns due to unanticipated findings or complications [12, 15]. These errors propagate through the hierarchical structure, potentially yielding suboptimal timing and redosing sequences [16, 19]. Off-policy evaluation estimates may be biased when target and behavior policies differ substantially or when unmeasured confounders—including surgical technique, hemostasis quality, and wound closure—influence SSI risk independently of prophylaxis [2, 7, 25, 26]. The assumption that tissue concentration mediates prophylaxis effectiveness may oversimplify the interplay of host immunity, bacterial inoculum, and wound environment determining SSI development [3, 11].

Clinical limitations

Clinical implementation faces barriers related to information technology integration, provider acceptance, and medicolegal considerations. Interoperable interfaces with scheduling and anesthesia information systems are not universally available, and computational infrastructure for real-time prediction may exceed institutional capabilities [17, 18]. Anesthesiologist acceptance cannot be assumed, particularly when recommendations conflict with experiential knowledge or lack transparent rationale [14, 22]. Medicolegal implications of AI-recommended redosing remain uncertain regarding liability for adverse drug events or SSIs following withheld recommendations [28, 29]. Performance may vary across institutions with different populations and resistance patterns, necessitating local validation requiring investment in data infrastructure and clinical informatics expertise [4, 9].

Conclusion

This article has presented a conceptual framework for applying hierarchical reinforcement learning to personalized perioperative antibiotic prophylaxis, addressing optimal preoperative timing and adaptive intraoperative redosing. The framework decomposes the problem into two policy levels operating on distinct timescales, with a high-level manager determining administration time and a low-level worker managing redosing, informed by uncertainty-aware procedure duration predictions.

The key advantages include multi-timescale optimization aligned with perioperative care structure, integration of duration prediction uncertainty into risk-sensitive planning, and personalization based on patient-specific infection risk factors. By simultaneously optimizing SSI prevention and antibiotic stewardship through a carefully designed reward function, the framework aims to achieve improvements unattainable through static guidelines.

Significant limitations include irreducible duration prediction uncertainty for outlier cases, potential off-policy evaluation bias from unmeasured confounding, and substantial implementation challenges related to electronic health record integration and provider acceptance. These define the boundaries for performance assessment and cautions for any clinical deployment.

The pathway forward involves evaluation on surgical databases such as the National Surgical Quality Improvement Program or single-institution electronic health

record datasets. Prospective silent evaluation would enable comparison of policy recommendations against actual practice without confounding from provider response to decision support. Such evaluation, followed by phased clinical implementation with safety and efficacy monitoring, represents the necessary next step toward realizing reinforcement learning's potential to personalize perioperative antibiotic prophylaxis.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 21 May 2025 Revised: 06 Jul 2025 Accepted: 07 Aug 2025
Published online: 20 January 2026

References

Berríos-Torres SI, Umscheid CA, Bratzler DW, Leas B, Stone EC, Kelz RR, et al. Centers for disease control and prevention guideline for the prevention of surgical site infection, 2017. JAMA Surg. 2017;152(8):784-91.

de Jonge SW, Gans SL, Atema JJ, Solomkin JS, Dellinger PE, Boermeester MA. Timing of preoperative antibiotic prophylaxis in 54,552 patients and the risk of surgical site infection: a systematic review and meta-analysis. Medicine (Baltimore). 2017;96(29):e6903.

Weber WP, Mujagic E, Zwahlen M, Bundi M, Hoffmann H, Soysal SD, et al. Timing of surgical antimicrobial prophylaxis: a phase 3 randomised controlled trial. Lancet Infect Dis. 2017;17(6):605-14.

Branch-Elliman W, O'Brien W, Strymish J, Itani K, Wyatt C, Gupta K. Association of duration and type of surgical prophylaxis with antimicrobial-associated adverse events. JAMA Surg. 2019;154(7):590-8.

de Jonge SW, Boldingh QJ, Solomkin JS, Dellinger EP, Egger M, Salanti G, et al. Effect of postoperative continuation of antibiotic prophylaxis on the incidence of surgical site infection: a systematic review and meta-analysis. Lancet Infect Dis. 2020;20(10):1182-92.

de Jonge SW, Boldingh QJ, Koch AH, Daniels L, de Vries EN, Spijkerman IJ, et al. Timing of preoperative antibiotic prophylaxis and surgical site infection: TAPAS, an observational cohort study. Ann Surg. 2021;274(4):e308-e314.

Bertschi D, Weber WP, Zeindler J, Stekhoven D, Mechera R, Salm L, et al. Antimicrobial prophylaxis redosing reduces surgical site infection risk in prolonged duration surgery irrespective of its timing. World J Surg. 2019;43(10):2420-5.

O'Hara LM, Thom KA, Preas MA. Update to the Centers for Disease Control and Prevention and the Healthcare Infection Control Practices Advisory Committee Guideline for the Prevention of Surgical Site Infection (2017): a summary, review, and strategies for implementation. Am J Infect Control. 2018;46(6):602-9.

Sartelli M, Labricciosa FM, Casini B, Cortese F, Cricca M, Facciolà A, et al. Optimizing surgical antibiotic prophylaxis in the era of antimicrobial resistance: a position paper from the Italian Multidisciplinary Society for the Prevention of Healthcare-Associated Infections (SIMPIOS). Pathogens. 2025;14(10):1031.

Al Mamlook RE, Wells LJ, Sawyer R. Machine-learning models for predicting surgical site infections using patient pre-operative risk and surgical procedure factors. Am J Infect Control. 2023;51(5):544-50.

Chen KA, Joisa CU, Stem JM, Guillem JG, Gomez SM, Kapadia MR. Improved prediction of surgical-site infection after colorectal surgery using machine learning. Dis Colon Rectum. 2023;66(3):458-66.

Xiong C, Zhao R, Xu J, Liang H, Zhang C, Zhao Z, et al. Construct and validate a predictive model for surgical site infection after posterior lumbar interbody fusion based on machine learning algorithm. Comput Math Methods Med. 2022;2022:2697841.

McLean KA, Sgrò A, Brown LR, Buijs LF, Mountain KE, Shaw CA, et al. Multimodal machine learning to predict surgical site infection with healthcare workload impact assessment. NPJ Digit Med. 2025;8(1):121.

Muaddi H, Choudhary A, Lee F, Anderson SS, Habermann E, Etzioni D, et al. Imaging-based surgical site infection detection using artificial intelligence. Ann Surg. 2025;282(3):419-28.

Elhage SA, Deerenberg EB, Ayuso SA, Murphy KJ, Shao JM, Kercher KW, et al. Development and validation of image-based deep learning models to predict surgical complexity and complications in abdominal wall reconstruction. JAMA Surg. 2021;156(10):933-40.

Bartek MA, Saxena RC, Solomon S, Fong CT, Behara LD, Venigandla R, et al. Improving operating room efficiency: machine learning approach to predict case-time duration. J Am Coll Surg. 2019;229(4):346-54.

Strömblad CT, Baxter-King RG, Meisami A, Yee SJ, Levine MR, Ostrovsky A, et al. Effect of a predictive model on planned surgical duration accuracy, patient wait time, and use of presurgical resources: a randomized clinical trial. JAMA Surg. 2021;156(4):315-21.

Jiao Y, Xue B, Lu C, Avidan MS, Kannampallil T. Continuous real-time prediction of surgical case duration using a modular artificial neural network. Br J Anaesth. 2022;128(5):829-37.

Spence C, Shah OA, Cebula A, Tucker K, Sochart D, Kader D, et al. Machine learning models to predict surgical case duration compared to current industry standards: scoping review. BJS Open. 2023;7(6):zrad113.

Riahi V, Hassanzadeh H, Khanna S, Boyle J, Syed F, Biki B, et al. Improving preoperative prediction of surgery duration. BMC Health Serv Res. 2023;23(1):1343.

Rozario D. Can machine learning optimize the efficiency of the operating room in the era of COVID-19? Can J Surg. 2020;63(6):E527.

Lex JR, Abbas A, Mosseri J, Toor JS, Simone M, Ravi B, et al. Using machine learning to predict-then-optimize elective orthopedic surgery scheduling to improve operating room utilization: retrospective study. JMIR Med Inform. 2025;13(1):e70857.

Zhong C, Liao K, Chen W, Liu Q, Peng B, Huang X, et al. Hierarchical reinforcement learning for automatic disease diagnosis. Bioinformatics. 2022;38(16):3995-4001.

Du X, Chen H, Yang B, Long C, Zhao S. HRL4EC: hierarchical reinforcement learning for multi-mode epidemic control. Inf Sci. 2023;640:119065.

Tang S, Makar M, Sjoding M, Doshi-Velez F, Wiens J. Leveraging factored action spaces for efficient offline reinforcement learning in healthcare. Adv Neural Inf Process Syst. 2022;35:34272-86.

Tang S, Wiens J. Model selection for offline reinforcement learning: practical considerations for healthcare settings. In: Proceedings of Machine Learning for Healthcare Conference. PMLR; 2021. p. 2-35.

Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med. 2018;24(11):1716-20.

Roggeveen L, El Hassouni A, Ahrendt J, Guo T, Fleuren L, Thoral P, et al. Transatlantic transferability of a new reinforcement learning model for optimizing haemodynamic treatment for critically ill patients with sepsis. Artif Intell Med. 2021;112:102003.

Wang Y, Liu A, Yang J, Wang L, Xiong N, Cheng Y, et al. Clinical knowledge-guided deep reinforcement learning for sepsis antibiotic dosing recommendations. Artif Intell Med. 2024;150:102811.