ORIGINAL RESEARCH **Open access**

Reinforcement Learning for Intravenous Fluid Resuscitation in Septic Shock: A Position Paper on Safety Constraints, Reward Design, and Clinical Oversight

Claire Martin[1], Julien Robert[2*], Sophie Bernard[1], Antoine Girard[2]

Abstract

Septic shock, defined as sepsis with persistent hypotension despite adequate fluid resuscitation and requiring vasopressors, has a mortality rate of 30–50% despite modern treatment. Intravenous fluids remain the cornerstone of early therapy, with guidelines recommending at least 30 mL/kg of crystalloids within the first three hours. However, both insufficient and excessive fluid administration can be harmful, making individualized, data-driven management essential. Reinforcement learning (RL) has been proposed to optimize fluid and vasopressor dosing in sepsis using retrospective ICU data. While models such as the AI Clinician suggest potential survival benefits, they often prioritize long-term outcomes like mortality and overlook short-term harms such as fluid overload and organ injury, raising safety concerns. Safety constraints and harm-aware reward design are essential in RL systems for septic shock. Pure outcome optimization is insufficient, and clinical AI must include mechanisms to prevent unsafe actions and ensure adherence to safety limits. Offline RL is vulnerable to distributional shift and unsafe extrapolation. Reward functions focused only on survival ignore acute complications, leading to unsafe policies. Human-in-the-loop oversight is necessary to maintain clinical accountability and enable intervention. RL systems should include action constraints, conservative learning with uncertainty estimation, and reward penalties for fluid overload indicators. Regulatory bodies and journals should require safety validation, and clinicians must retain override authority and transparency in decision-making. RL in septic shock management must prioritize patient safety through constraints, harm-aware rewards, and clinical oversight. Without these safeguards, deployment risks patient harm and loss of trust in clinical AI.

Keywords Reinforcement learning, Safety constraints, Septic shock, Fluid resuscitation, Reward design, Human-in-the-loop

*Correspondence:
Julien Robert
julien.robert@gmail.com
[1] Department of Healthcare Informatics and AI, University of Lyon, Lyon, France
[2] Department of Intelligent Clinical Systems, University of Strasbourg, Strasbourg, France

Introduction

Septic shock is a life-threatening condition characterized by sepsis-induced hypotension that persists despite adequate fluid resuscitation and requires vasopressors to maintain a mean arterial pressure of at least 65 mmHg [1]. Mortality remains high at 30-50%, underscoring the urgency of optimizing early interventions. Intravenous fluid resuscitation serves as the first-line therapy, yet both insufficient volume leading to ongoing hypoperfusion and excessive administration causing fluid overload contribute to worse outcomes [2]. This delicate balance highlights the limitations of one-size-fits-all protocols in a heterogeneous patient population.

Reinforcement learning has been proposed as a data-driven approach to personalize fluid dosing in sepsis management [3, 4]. Komorowski et al. demonstrated that

an RL agent could learn treatment strategies associated with improved outcomes in retrospective data [5]. However, subsequent analyses and related work raised serious concerns regarding the safety of such policies when extrapolated to real-world settings [6, 7]. These models often operate in offline modes due to the ethical impossibility of online experimentation in critically ill patients, yet they frequently overlook critical safety boundaries [8].

We argue that current RL approaches for fluid resuscitation in septic shock are unsafe for clinical deployment because they: (1) lack adequate safety constraints to prevent harmful actions, (2) use reward functions that ignore immediate treatment harms such as pulmonary edema and acute kidney injury, and (3) bypass meaningful clinical oversight. We propose specific design requirements centered on constrained action spaces, harm-penalizing rewards, and human-in-the-loop architectures to address each of these failures. Only through these measures can RL move beyond retrospective promise to prospective safety.

Figure 1 illustrates a safety-constrained reinforcement learning architecture that integrates action masking, harm-aware reward design, and human-in-the-loop oversight to ensure clinically safe fluid resuscitation decisions.

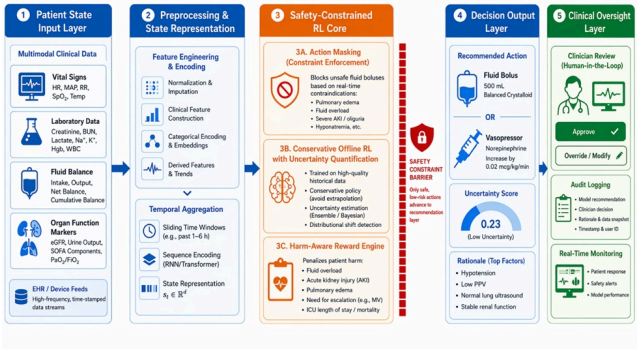

Figure 1. Safety-Constrained Reinforcement Learning Architecture for Intravenous Fluid Resuscitation in Septic Shock

Septic Shock and Fluid Resuscitation

Clinical guidelines and uncertainty

The Surviving Sepsis Campaign 2021 guidelines recommend administering at least 30 mL/kg of intravenous crystalloid fluid within the first three hours for patients with sepsis-induced hypoperfusion or septic shock [1]. They further suggest using balanced crystalloids rather than normal saline and advocate for dynamic measures to guide ongoing resuscitation over static parameters alone [1, 2]. Despite these recommendations, significant uncertainty persists regarding the optimal volume, timing, and type of fluid for individual patients, as responses vary widely based on comorbidities, capillary leak, and evolving organ function.

We contend that these guidelines, while evidence-based, leave substantial room for personalization that static protocols cannot address. RL offers a theoretical avenue to learn patient-specific policies from large observational datasets [3, 5]. Yet without embedding guideline-derived safety rules directly into the learning process, agents risk violating fundamental clinical principles [8]. The evidence gaps in fluid management demand cautious, constrained innovation rather than unconstrained optimization.

Harms of over-resuscitation

Fluid overload in septic shock is strongly associated with pulmonary edema, prolonged mechanical ventilation, and increased mortality. Observational data consistently link positive fluid balance and higher percentages of fluid overload to worse respiratory outcomes, including reduced ventilator-free days and heightened risk of acute respiratory distress. These complications arise because excess extravascular lung water impairs gas exchange and increases work of breathing, often necessitating escalated support.

Moreover, fluid overload contributes to acute kidney injury through increased renal venous pressure, interstitial edema, and compartment effects that reduce glomerular filtration [2]. Studies demonstrate that patients developing significant fluid accumulation face higher rates of organ dysfunction and death, even after adjusting for illness severity. We argue that any RL system ignoring these well-documented short-term harms in favor of long-term survival metrics will systematically recommend policies that trade immediate patient safety for speculative future gains [5, 9].

The Promise and Peril of RL for Fluid Management

Potential benefits of RL

Reinforcement learning holds genuine promise for personalizing intravenous fluid resuscitation by learning sequential policies that adapt continuously to a patient's evolving physiology using large ICU datasets [3, 4]. Unlike traditional protocols, RL can theoretically discover nuanced strategies that balance fluid administration with vasopressor needs while accounting for individual variability in sepsis response [5]. This data-driven personalization could reduce both under- and over-resuscitation compared with rigid guideline thresholds.

We contend that when properly constrained, RL could integrate multimodal data, including vital signs, laboratory trends, and fluid balance metrics, to support more precise decisions [6]. Early studies have shown agents capable of identifying treatment patterns associated with better aggregate outcomes in retrospective cohorts [5]. The adaptive nature of RL aligns well with the dynamic, time-sensitive nature of septic shock management.

Documented failures and risks

Despite this potential, documented failures in RL applications for sepsis reveal serious risks, particularly around distributional shift in offline settings and extrapolation beyond observed clinician behaviors [8, 10]. The AI Clinician and similar models have faced criticism for policies that may not generalize safely, with reanalyses highlighting concerns over unaddressed harms [5, 7]. Offline RL, while necessary for ethical reasons, struggles with conservative estimation of rarely observed actions that could prove dangerous in practice [8].

We argue that these issues stem from fundamental design choices rather than data limitations alone. Without explicit safety mechanisms, agents can learn policies that appear optimal under retrospective evaluation but recommend excessive fluid volumes or ignore contraindications such as existing edema [9]. The peril lies not in RL itself but in its unconstrained application to a domain where errors carry immediate life-threatening consequences.

Table 1 delineates the fundamental failure modes of reinforcement learning in septic shock fluid management and maps each to necessary safety-critical design interventions.

Table 1. Structural Failure Modes in Reinforcement Learning for Fluid Resuscitation and Corresponding Safety-Critical Design Requirements

Domain of Failure	Underlying Technical Issue	Clinical Consequence	Wh Standar Fail
Action Selection	Unconstrained action space	Excessive fluid boluses in contraindicated states	RL optir over mathema valid ac regardle clinic feasib
Offline Learning	Distributional shift and extrapolation error	Unsafe recommendations in rare or unseen states	Q-val overestir for out distribu actio
Reward Design	Sparse, delayed outcome signals	Ignoring fluid overload and organ injury	Cred assignr proble obscu short-t harm
Policy Evaluation	Reliance on off-policy evaluation	Undetected catastrophic edge-case failures	Counterf estima biased in dimens settin
Deployment Context	Lack of clinician oversight	Automation bias and delayed correction	RL treat autonor rather assist syste

Safety Constraints

Action masking

Action masking offers a practical mechanism to integrate clinical rules directly into RL agents, preventing the selection of harmful fluid boluses in the presence of pulmonary edema, elevated filling pressures, or other contraindications [8]. By dynamically masking actions outside safe ranges derived from guidelines and patient-specific physiology, developers can enforce hard safety boundaries while still allowing optimization within permissible spaces. This approach transforms the action space from fully unconstrained to clinically grounded.

We contend that simple post-hoc filtering is insufficient; masking must occur during training and inference to shape policy learning itself [10]. Studies on safe RL in healthcare demonstrate that such constraints improve alignment with expert practices without sacrificing performance on feasible actions [8]. For fluid resuscitation, masking high-volume boluses in overload states is not optional but a minimum requirement for any deployable system.

Safe exploration vs offline RL

Online RL exploration is ethically prohibited in ICU settings, as it would require deliberately testing potentially harmful policies on vulnerable patients. Consequently, the field relies on offline or batch RL trained on historical data, yet these methods face challenges from distributional shift and poor coverage of rare but critical states [8, 10]. Conservative approaches, such as uncertainty-penalized or constrained Q-learning, help mitigate overestimation of out-of-distribution actions but still require careful validation.

We argue that offline RL alone cannot guarantee safety without additional safeguards like uncertainty quantification and conservative policy iteration [8]. Work on safe offline RL for sepsis treatment underscores that standard methods may still propose risky interventions not frequently observed in training data. Safe exploration must therefore be simulated through constrained offline techniques rather than assumed through data volume.

Position statement on safety

We contend that any clinical RL agent for intravenous fluid resuscitation in septic shock must be provably safe within a clinically defined action space before any form of human testing or deployment [8, 10]. Provable safety here means explicit enforcement of hard constraints derived from guidelines, such as maximum fluid volumes conditional on organ function markers, combined with uncertainty-aware policies that defer or flag low-confidence recommendations. Retrospective performance metrics alone provide false reassurance when safety violations remain possible.

Failure to implement such constraints represents negligence rather than technical limitation. The ICU environment demands that RL systems respect the same "do no harm" imperative as clinicians. We insist that safety must be designed in from the outset, not audited after the fact.

Reward Design

Problems with outcome-only rewards

Reward functions based solely on long-term outcomes such as 90-day survival suffer from critical limitations in sparse, delayed feedback typical of sepsis trajectories [5, 9]. These designs create difficult credit assignment problems, where the agent struggles to link early fluid decisions to distant survival while ignoring intermediate harms like progressive fluid overload. Consequently, policies may optimize for survival at the expense of increased short-term organ stress.

We argue that purely outcome-driven rewards encourage myopic or overly aggressive strategies because they fail to penalize clinically evident intermediate harms [5]. Work on RL for sepsis has shown that such rewards can lead to policies differing markedly from clinician behavior without clear safety guarantees [9]. This mismatch highlights a fundamental flaw: survival alone does not equate to acceptable treatment quality when harm pathways are ignored.

Incorporating harm into reward

Effective reward design must incorporate explicit penalties for markers of fluid overload, including positive fluid balance thresholds, rising extravascular lung water indicators, worsening oxygenation, or acute kidney injury progression [9, 11]. Composite rewards can balance survival and organ function preservation by assigning negative values to events such as pulmonary edema or prolonged ventilation needs while rewarding stable hemodynamics and lactate clearance. This harm-aware approach better aligns the objective with clinical priorities of minimizing iatrogenic injury.

Table 2 contrasts outcome-only and harm-aware reward formulations, demonstrating why incorporating explicit penalties for fluid-related harm is essential for clinically valid reinforcement learning.

Table 2. Comparative Framework for Reward Function Design: Outcome-Only vs Harm-Aware Approaches in Septic Shock RL

Dimension	Outcome-Only Reward Design	Harm-Aware Composite Reward Design	Theoretical Implication
Objective Signal	Long-term survival (e.g., 90-day mortality)	Survival + intermediate harm penalties (fluid overload, AKI, edema)	Sparse vs dense reward structure
Temporal Sensitivity	Delayed feedback	Immediate and longitudinal feedback	Weak credit assignment v improved temporal attribution
Interpretability	Low (black-box survival optimization)	Higher (clinically interpretable components)	Opaque optimization vs structured objective decompositio
Policy Behavior	Potentially aggressive or myopic	Balanced, conservative decision-making	Over-optimization risk vs constrained optimization
Ethical Validity	Ignores iatrogenic harm pathways	Explicitly encodes “do no harm” principle	Misaligned objective vs ethically grounded optimization

We contend that dense, clinically meaningful intermediate rewards improve learning stability and produce policies that clinicians can more readily trust. By engineering rewards that encode both benefits and harms, RL systems can learn to avoid the very complications that undermine long-term outcomes [9]. Such designs move beyond simplistic survival maximization toward holistic patient benefit.

Position statement on reward

We argue that reward functions must explicitly encode both benefits and harms of intravenous fluid treatment. A reward that ignores harm will inevitably recommend harmful policies, as the optimization process has no incentive to avoid them [5, 9]. Harm encoding is not an optional refinement but a core ethical requirement for any RL system claiming clinical relevance in septic shock management.

Only through deliberate, transparent reward engineering that reflects real-world trade-offs can the field produce agents worthy of consideration for bedside use. We reject outcome-only formulations as clinically inadequate and demand that future work prioritize harm-aware designs validated against expert judgment.

Clinical Oversight

Human-in-the-loop requirements

Human-in-the-loop configurations are essential for any reinforcement learning system applied to intravenous fluid resuscitation in septic shock, ensuring that RL serves strictly as a decision support tool rather than an autonomous agent [12]. Clinicians must retain full veto power over recommendations, with the system providing transparent explanations for suggested fluid volumes or withholding actions. Real-time monitoring of patient physiology must trigger immediate alerts when RL proposals approach or violate predefined safety thresholds [12].

We contend that delegating fluid management decisions entirely to RL would constitute an unacceptable abdication of clinical responsibility. The dynamic and uncertain nature of septic shock demands continuous human judgment that no current model can fully replicate [5]. Oversight mechanisms must therefore embed clinician confirmation for all high-impact actions and maintain audit trails for post hoc review.

Position statement on oversight

We contend that RL for fluid resuscitation must operate in a human-in-the-loop configuration with real-time monitoring, override capability, and mandatory clinician confirmation of all recommendations above a safety threshold [12]. This architecture preserves accountability while allowing the system to learn from expert overrides and refine future suggestions. Without such safeguards, even well-constrained models risk introducing systematic errors that clinicians may not immediately recognize [5, 8].

We argue that meaningful oversight is non-negotiable for ethical deployment. RL should augment rather than replace clinical expertise, particularly in a domain where small deviations in fluid balance can precipitate life-threatening complications. Human-in-the-loop designs represent the only responsible pathway toward eventual integration in critical care.

Counterarguments Addressed

"Offline RL can be evaluated with off-policy evaluation"

Proponents often claim that offline RL policies can be safely evaluated using off-policy evaluation techniques such as counterfactual estimators [13]. However, these methods carry well-known biases, particularly in high-dimensional state spaces with limited action coverage typical of ICU datasets [8, 10]. Counterfactual evaluation cannot reliably detect rare but catastrophic safety violations that fall outside the observed data distribution.

We argue that reliance on off-policy evaluation provides false confidence rather than genuine safety guarantees. Distributional shift and extrapolation errors remain fundamental challenges that no estimator fully resolves in this context [8]. We reject the notion that technical evaluation metrics alone suffice for clinical readiness when patient lives are at stake.

"We can start with low-risk patients"

Some suggest initiating RL deployment in lower-risk septic patients to gather prospective data before broader application. Yet no patient in septic shock qualifies as truly low-risk for fluid-related harm, given the unpredictable progression of capillary leak, organ dysfunction, and comorbidities [2]. Even modest fluid overload can rapidly escalate respiratory or renal failure in seemingly stable individuals.

We contend that stepwise testing on “low-risk” subgroups still exposes vulnerable patients to unproven policies without adequate safeguards [1]. The heterogeneity of septic shock makes risk stratification unreliable for experimental deployment. Any introduction of RL must incorporate conservative constraints and human oversight from the first prospective use, rather than assuming certain subgroups are expendable for validation.

"Clinicians already make errors — RL may do better"

Critics sometimes argue that because clinicians commit errors in fluid management, RL systems should be permitted similar or even higher error rates if they demonstrate aggregate outcome improvements [5]. This lower-bar fallacy ignores the fundamentally different risk profile of algorithmic versus human mistakes, where RL errors tend to be systematic, correlated across similar patients, and difficult to detect in real time. Human errors, while frequent, benefit from contextual judgment and immediate corrective action that models lack.

We argue that introducing RL does not absolve the field from demanding higher safety standards than current practice. The goal must be to reduce harm overall, not merely match or slightly exceed imperfect human performance [9]. Systematic algorithmic failures could erode trust far more severely than sporadic clinician variability, making safety constraints and oversight even more critical.

Recommendations

For researchers

Researchers developing RL for fluid resuscitation must prioritize reporting explicit safety metrics alongside performance outcomes, including the frequency of masked or overridden harmful actions and uncertainty estimates for out-of-distribution states [8, 10]. They should design harm-aware reward functions that explicitly penalize fluid overload events such as pulmonary edema and acute kidney injury rather than relying solely on long-term survival [9, 11]. Publishing failure cases and ablation studies on safety constraints should become standard practice to accelerate collective progress.

We contend that transparency in safety limitations is as important as reporting performance gains. Only by openly documenting where current approaches fail can the community build more robust systems. Researchers must move beyond optimistic retrospective results and embrace rigorous safety engineering as a core scientific responsibility [5, 8].

For journal editors and reviewers

Journal editors and reviewers should reject manuscripts on clinical RL for sepsis that do not include explicit safety constraints, action masking details, or harm-penalizing reward components [8, 9]. Submissions must demonstrate how the proposed agent respects clinical guidelines on fluid resuscitation and quantifies risk of fluid overload [1, 2]. Reviewers should demand clear reporting of counterfactual safety violations and human-in-the-loop compatibility.

We argue that lowering standards for AI papers in high-stakes medicine perpetuates unsafe practices. Journals have a gatekeeping duty to protect patients by enforcing minimum safety criteria before publication. Without this discipline, the literature will continue to overstate readiness and understate risks of RL deployment in critical care.

For regulatory bodies (FDA, EMA)

Regulatory bodies such as the FDA and EMA must require safety certification for any RL-based system intended for fluid management in septic shock, including formal verification of constrained action spaces and worst-case harm analysis [8, 10]. Human-in-the-loop mandates and post-market surveillance protocols should be compulsory, with mandatory clinician override functionality and real-time safety monitoring. Approval pathways must include phased validation from retrospective analysis through silent prospective testing before any interventional use.

We contend that treating clinical RL as standard software without rigorous oversight ignores the unique risks of autonomous decision influence in life-threatening conditions. Regulators must establish clear guidelines that prioritize harm prevention over innovation speed. Failure to do so risks widespread adoption of systems that could cause preventable harm at scale.

For clinicians and hospital administrators

Clinicians and hospital administrators should demand robust evidence of safety constraints and harm-aware reward design before considering any RL tool for fluid resuscitation [8, 9]. Systems must include transparent explanations, real-time override capabilities, and audit logs for every recommendation. Deployment should only occur within tightly controlled human-in-the-loop frameworks with mandatory training on system limitations.

We argue that passive acceptance of vendor claims without independent verification endangers patients and exposes institutions to liability. Administrators must insist on independent safety audits and phased introduction protocols. Clinicians remain ultimately responsible for patient care and must retain meaningful control over any AI-assisted decisions.

Path Forward

Safe RL research priorities

Safe RL research priorities should focus on conservative offline RL methods that incorporate uncertainty quantification and explicit action masking derived from Surviving Sepsis Campaign guidelines [1, 8, 10]. Reward function engineering must move beyond sparse survival objectives toward dense, clinically interpretable signals that penalize fluid overload and organ dysfunction while rewarding hemodynamic stability [9, 11]. Development of standardized benchmarks for safety violations in fluid resuscitation would enable fair comparison across approaches.

We contend that these priorities represent the minimum technical foundation for credible progress. Incremental improvements in predictive accuracy without corresponding safety advances are insufficient. The community must treat safety as a first-class research objective rather than an afterthought.

Validation pathway

A responsible validation pathway must progress stepwise from retrospective counterfactual evaluation through silent prospective monitoring to controlled human-in-the-loop trials before any randomized interventional study [8, 13]. Each stage should include predefined safety stopping rules based on fluid overload events and organ injury markers [2]. Only after demonstrating consistent adherence to clinical safety boundaries in prior phases should broader deployment be considered.

We argue that shortcuts in validation expose patients to unnecessary risk and undermine long-term trust in clinical AI. This structured pathway balances innovation with prudence. Rigorous, staged evaluation is the only ethical route to eventual safe integration of RL in septic shock management.

Conclusion

Reinforcement learning for intravenous fluid resuscitation in septic shock remains promising yet currently unsafe for clinical deployment due to inadequate safety constraints, flawed reward designs, and insufficient clinical oversight. Prominent studies have highlighted the potential of data-driven personalization, but retrospective optimism has overshadowed critical risks of fluid overload and organ harm. Without fundamental redesign, RL systems risk recommending policies that prioritize speculative long-term gains over immediate patient safety.

We contend that safety constraints, harm-aware reward functions, and mandatory human-in-the-loop architectures are non-negotiable requirements rather than optional enhancements. These elements must be embedded at every stage of development, evaluation, and potential deployment. Anything less violates the core ethical duty to “do no harm” in critical care.

The consequences of ignoring these requirements are severe: preventable patient harm, erosion of clinician trust in AI, and potential regulatory backlash that could stall legitimate innovation for years. Unsafe deployment would not only damage individual patients but also set back the entire field of AI for healthcare.

We call on researchers, journals, regulators, and clinicians to adopt a strict safety-first design philosophy immediately. Only by enforcing these standards can RL evolve from an experimental curiosity into a trustworthy tool for improving outcomes in septic shock. The time for unconstrained exploration has passed; responsible, constrained, and overseen development is now mandatory.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 14 May 2021 Revised: 09 Jul 2021 Accepted: 25 Aug 2021
Published online: 20 January 2022

References

Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Crit Care Med. 2021;49(11):e1063-e1143. https://doi.org/10.1097/CCM.0000000000005337.

Macdonald S. Fluid resuscitation in patients presenting with sepsis: current insights. Open Access Emerg Med. 2022;14:633-8. https://doi.org/10.2147/OAEM.S319777.

Jia Y, Burden J, Lawton T, Habli I. Safe reinforcement learning for sepsis treatment. In: 2020 IEEE Int Conf Healthc Inform. 2020. p. 1-7. https://doi.org/10.1109/ICHI48887.2020.9374387.

Raghu A, Komorowski M, Ahmed I, Celi LA, Szolovits P, Ghassemi M. Reinforcement learning for sepsis treatment:

baselines and analysis. In: Mach Learn Healthc Conf Proc. 2017;68:241-52.

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. https://doi.org/10.1038/s41591-018-0213-5.

Mollura M, Drudi C, Lehman LW, Barbieri R. A reinforcement learning application for optimal fluid and vasopressor interventions in septic ICU patients. In: 2022 44th Annu Int Conf IEEE Eng Med Biol Soc. 2022. p. 321-324. https://doi.org/10.1109/EMBC48229.2022.9870975.

Su L, Li Y, Liu S, Zhang S, Zhou X, Weng L, et al. Establishment and implementation of potential fluid therapy balance strategies for ICU sepsis patients based on reinforcement learning. Front Med (Lausanne). 2022;9:766447. https://doi.org/10.3389/fmed.2022.766447.

Liu R, Greenstein JL, Fackler JC, Bergmann J, Bembea MM, Winslow RL. Offline reinforcement learning with uncertainty for treatment strategies in sepsis. arXiv. 2021;2107.04491.

Nanayakkara T, Clermont G, Langmead CJ, Swigon D. Unifying cardiovascular modelling with deep reinforcement learning for uncertainty aware control of sepsis treatment. PLOS Digit Health. 2022;1(2):e0000012. https://doi.org/10.1371/journal.pdig.0000012.

Huang Y, Cao R, Rahmani A. Reinforcement learning for sepsis treatment: a continuous action space solution. In: Mach Learn Healthc Conf Proc. 2022;193:631-47.

Yu C, Ren G, Liu J. Deep inverse reinforcement learning for sepsis treatment. In: 2019 IEEE Int Conf Healthc Inform. 2019. p. 1-3. https://doi.org/10.1109/ICHI.2019.8904727.

Kim HI, Park S. Sepsis: early recognition and optimized treatment. Tuberc Respir Dis (Seoul). 2019;82(1):6-14. https://doi.org/10.4046/trd.2017.0041.

Oberst M, Sontag D. Counterfactual off-policy evaluation with gumbel-max structural causal models. In: Proc Int Conf Mach Learn. 2019;97:4881-90.