Leveraging AI-based prediction in perioperative and critical care: from model development to clinical implementation

Meijden, S.L. van der

Chapter 10

Summary

In this dissertation, the potential of Artificial Intelligence (AI) for predicting patient outcomes in perioperative and critical care was investigated. The chapters of this dissertation align with the AI development and evaluation trajectory, ranging from (data) preparations to implementing and governing AI. The main use case discussed throughout this dissertation was PERISCOPE, an AI tool predicting the risk of postoperative infection within 7- and 30-days of surgery.

Chapter 2 provides a review of the possibilities and challenges of AI and Machine Learning (ML) in perioperative care. The basic concepts of AI and ML are discussed as well as promising AI applications in anesthesia, surgery, complication prediction, operating room logistics, and nursing practice. The steps of model development and validation are outlined, including relevant regulations. Lastly, the challenges and pitfalls that physicians and nurses should be aware of before implementing AI tools in their clinical practice are discussed.

Phase 0: Preparations prior to model development
Chapter 3 covers a pre-implementation survey study among 64 intensive care unit (ICU) physicians to investigate their perspectives on AI and their current clinical decision-making behavior. This survey aimed to gather information on physicians' perspectives towards AI and current decision-making behavior before the potential implementation of an AI-based ICU readmission risk prediction model. Respondents showed high familiarity with AI and believed that AI could support them in their daily work. However, significant differences between the development and non-development sites show that it is important to establish an effective implementation process with a focus on training and informing end users.

Chapter 4 reviews the different methods used to identify patients with (postoperative) infections in prediction and surveillance studies. To train and validate postoperative infection prediction models, it must reliably be determined based on electronic health record (EHR) data which patients had an infection and which did not. In this study, we found that there are 75 different methods and definitions to identify patients with infections and that manual chart review is still the predominant method, which limits the scalability of AI tools.

Phase 1: AI model development
Chapter 5 provides a feasibility study of PERISCOPE with internal validation results and a validation of the label of postoperative infections compared to human chart review. The automated infection method to label patients for the outcome had a 93% (54/60) agreement compared to manual chart review. The area under the receiver operating characteristic curve (AUROC), indicating the discriminative performance of the model, was 0.81 (95% confidence interval (CI) 0.80-0.83) on the temporal test dataset (n=9,722 procedures). Based on these results, the PERISCOPE model was deemed ready for external validation.

Phase 2: Assessment of AI performance and reliability
Chapter 6 presents the external validation results and updating approach of PERISCOPE in two Dutch and one Belgian hospital. A total of 253,010 procedures were included in this study,

where we showed that discriminative performance, calibration, and clinical utility significantly improved after updating the model per site. A thorough subgroup analysis was conducted to assess biases across different demographics and patient groups. The methodology described in this chapter can be applied to other AI-based prediction models to ensure local validity by accounting for domain shifts.

Phase 3: Clinical testing of AI

Chapter 7 evaluates the predictive performance of PERISCOPE against that of 51 surgeons and surgeons in training in predicting postoperative infection risk. PERISCOPE outperformed human participants when they were unsure about their prediction (AUROC 0.80 vs. 0.68) and clinicians in training (AUROC 0.77 vs. 0.73), but sample sizes were small. Calibration properties were better for the PERISCOPE model compared to clinicians. These findings indicate that PERISCOPE's predictions may be especially valuable in these contexts and will also provide supportive confirmation in other situations.

Phase 4: Implementing and governing AI

Chapter 8 provides an overview of fairness evaluation methods and a simplified framework for fairness evaluation of AI-based prediction models. The field of AI fairness aims to ensure that healthcare disparities are not increased due to biased algorithms. Through two use cases, we demonstrate that different fairness metrics may be applicable depending on the intended use and the underlying ethical framework. This practical guidance is crucial for AI developers and assessors aiming to promote more equitable AI implementations by assessing model fairness and the impact of bias mitigation strategies.

In conclusion, the potential of AI in predicting perioperative and critical care patient outcomes is gradually being realized by addressing data challenges and implementing robust validation processes. The steps towards effective implementation were demonstrated for the PERISCOPE tool. However, more clinical evidence is needed to assess the real-life impact of AI-based prediction models on patient outcomes, costs, and health equity.