Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230134

Identifying and Predicting Postoperative Infections Based on Readily Available Electronic Health Record Data

Siri Lise van der MEIJDEN[a,b,1,2], Anna van BOEKEL[b], Laurens SCHINKELSHOEK[a,2], Harry van GOOR[c], Mark de BOER[b], Ewout STEYERBERG[b], Bart GEERTS[a,2] and Sesmu ARBOUS[b]

[a] *Healthplus.ai, Amsterdam, The Netherlands*
[b] *Leiden University Medical Center, Leiden, The Netherlands*
[c] *Radboud University Medical Center, Nijmegen, The Netherlands*

ORCiD ID: S.L. van der Meijden https://orcid.org/0000-0002-9343-0899, A.M. van Boekel https://orcid.org/0000-0003-0774-0995, L.J. Schinkelshoek https://orcid.org/0000-0002-3112-5231, H. van Goor https://orcid.org/0000-0003-0323-4876, E.W. Steyerberg https://orcid.org/0000-0002-7787-0122, B.F. Geerts https://orcid.org/0000-0002-0210-7202, M.S. Arbous https://orcid.org/0000-0001-5242-3257

Abstract. Identification of postoperative infections based on retrospective patient data is currently done using manual chart review. We used a validated, automated labelling method based on registrations and treatments to develop a high-quality prediction model (AUC 0.81) for postoperative infections.

Keywords. Artificial Intelligence, Prediction, Postoperative infections, Electronic Health Record

1. Introduction

Postoperative infections are common complications with a global reported incidence of 9.0% [1]. To allow the development of prediction models for clinical decision support, we need to identify which patients had a postoperative infection in retrospective data. This process is called labelling, and is often done by manual review of the Electronic Health Record (EHR) as complications are severely under-registered [2]. Automation of this labour-intensive process is needed to enhance surveillance and scalability of prediction models. We aimed to develop and validate a postoperative infection risk prediction model using a domain knowledge-based labelling method based on readily available, non-free text, EHR patient data.

[1] Corresponding author: Siri Lise van der MEIJDEN, E-mail: S.L.van_der_meijden@lumc.nl
[2] S.L. and L.J. are employees and B.F. is owner and shareholder of Healthplus.ai

2. Methods

Adult surgical patients admitted to the Leiden University Medical Center (LUMC) between 2012 and 2022 were included for model development. In consultation with clinicians and data scientists, a definition was determined to label all postoperative bacterial infections that required complication or diagnosis registration, pharmacological treatment (initiating at least 24h after surgery for a minimum duration of 72 hours, excluding prophylactic regimes) and/or surgical intervention to treat infections. The dataset was split into a development part (2012-2020) and a temporal test part (2021-2022). Thirty cases with an infection and 30 without an infection according to the labelling method were randomly selected from the test dataset. Two clinicians independently determined for each case whether a postoperative infection occurred based on medication prescriptions, procedure information and complication and diagnosis registrations extracted from the EHR.

After this label validation step, an XGBoost model was trained on the development dataset to predict any bacterial infection within 30 days of surgery according to the definition. Medication prescriptions, patient characteristics, vital functions, comorbidities, and procedure characteristics were used as input features.

3. Results, Discussion and Conclusions

This research was done under the General Data Protection Regulation and a waiver for medical ethical approval from the LUMC was obtained (G18.129). Average overall infection rates were 12% (n=7,093 out of 59,106 procedures) in the development dataset and 13% (n=1,264 out of 9,722 procedures) in the test dataset.The automated infection label had 93% (54/60) agreement compared to manual labelling. We were able to predict postoperative infections on the test dataset with an area under the receiver operating characteristic curve of 0.81 (95% confidence interval (CI) 0.80-0.83), calibration slope of 0.81 (95% CI 0.77-0.86), and a positive net benefit, a measure that indicates clinical usefulness, for a broad range of threshold probabilities (0-85%).

The development of high-quality prediction models for clinical decision support requires reliable automatization of data processing and labelling. A treatment-based approach is needed to identify infections, since a definition based on solely registered complications would under estimate the risk. The next steps are to further validate the infection labelling method, improve the prediction model by adding more features to the dataset and externally validate the model on different hospital datasets to allow safe and broader implementation across hospital systems.

It is feasible to predict postoperative infections within 30 days of surgery with acceptable performance using an automated infection labelling method based on registrations and treatments of infection.

References

[1] International Surgical Outcomes Study group. Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries. Br J Anaesth. 2016 Oct 31;117(5):601-609.
[2] Ubbink DT, Visser A, Gouma DJ, Goslings JC. Registration of surgical adverse outcomes: a reliability study in a university hospital. BMJ Open. 2012 May 25;2(3):e000891.