International Journal of Population Data Science

Journal Website: www.ijpds.org

# Accuracy of EMR-based algorithms for Adverse Event Detection

Danielle Southern[1], Cheligeer Cheligeer[2], Guosong Wu[3], Cathy Eastwood[1], Yong Zeng[4], Bastien Boussat[5], Ward Flemons[6], Alan Forster[7], Yuan Xu[8], and Hude Quan[9]

[1]Centre for Health Informatics, Cumming School of Medicine, University of Calgary, Calgary, Canada

[2]Alberta Health Services, Calgary, Canada

[3]Department of Healthcare Analytics, Cape Breton University, Sydney, Canada

[4]Concordia Institute for Information Systems Engineering, Gina Cody School of Engineering and Computer Science, Concordia University, Montreal, Canada

[5]Clinical Epidemiology and Quality of Care Unit, University Grenoble Alpes, Faculty of Medicine, Grenoble University Hospital, Grenoble, France

[6]Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, Canada

[7]Department of Medicine, Faculty of Medicine and Health Sciences, McGill University, Montreal, Canada

[8]Department of Oncology, Cumming School of Medicine, University of Calgary, Calgary, Canada; Department of Surgery, Cumming School of Medicine, University of Calgary, Calgary, Canada

[9]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Canada; Centre for Health Informatics, Cumming School of Medicine, University of Calgary, Calgary, Canada

## Background

Adverse events (AEs) are often measured using administrative data, which has validity limitations. Electronic Medical Records (EMRs) provide rich, unstructured data that can improve AE detection through data science methods. Objective: To evaluate the accuracy of EMR-based algorithms for AE detection and estimate AE prevalence.

## Methods

We conducted a retrospective chart review validation study at a tertiary-care hospital (population ∼1.5M). Adult patients admitted between 2017–2022 with discharge summaries or narrative EMRs were included; only the most recent admission was analyzed. Seventeen algorithms were developed for specific AE categories and an overall "any AE" category using natural language processing (NLP) and machine learning. Prevalence, sensitivity, specificity, PPV, NPV, and model performance (AUC, accuracy, F1 score) were calculated using chart review as a reference.

## Results

Of 10,939 charts, 10,659 were linked to discharge data; 2,069 patients had ≥1 AE. AE prevalence ranged from 0.2%–6.0%. Sensitivity and specificity varied (57%–100%, 74%–96%). For "any AE," sensitivity was 63%, specificity 94%, PPV 72%, NPV 91%. Overall algorithm performance: AUC 78.4% (95% CI, 76.0–80.8), accuracy 87.6%, F1 score 67.2.

## Conclusion

NLP-based algorithms applied to EMRs show strong performance despite low AE prevalence, offering a scalable approach to enhance AE surveillance and patient safety.

https://doi.org/10.23889/ijpds.v11i5.3674