

Validation of a major and clinically relevant non-major bleeding phenotyping algorithm on electronic health records

Aaron Jun Yi Yap¹, Desmond Teo¹, Pei San ANG¹, Eng Soo Yap², Siew Har Tan¹, Celine Wei Ping Loke¹, and Sreemane Dorajoo¹

¹Health Sciences Authority

²National University Hospital

April 08, 2024

Abstract

Background: Bleeding is an important health outcome of interest in epidemiological studies. We aimed to develop and validate rule-based algorithms to identify major bleeding and all bleeding within real-world electronic healthcare data. **Methods:** We took a random sample (n=1630) of patient admissions to Singapore public hospitals in 2019 and 2020, stratifying by hospital and year of admission. We adopted the International Society on Thrombosis and Haemostasis definition for major bleeding. Presence of major bleeding and all bleeding was ascertained by two annotators through chart review. A total of 630 and 1,000 records were used for algorithm development and validation, respectively. We formulated two algorithms: sensitivity- and positive predictive value (PPV)-optimized algorithms. A combination of hemoglobin test patterns and diagnosis codes were used in the final algorithms. **Results:** During validation, diagnosis codes alone yielded low sensitivities for major bleeding (0.14) and all bleeding (0.24), although specificities and PPV were high (>0.97). For major bleeding, the sensitivity-optimized algorithm had much higher sensitivity and negative predictive values (NPV) (sensitivity=0.94, NPV=1.00), however false positive rates were also relatively high (specificity=0.90, PPV=0.34). PPV-optimized algorithm had improved specificity and PPV (specificity=0.96, PPV=0.52), with little reduction in sensitivity and NPV (sensitivity=0.88, NPV=0.99). For all bleeding events, our algorithms had less optimal performances, with lower sensitivities (0.53 to 0.61). **Conclusions:** The use of diagnosis codes alone misses many genuine major bleeding events. We have developed major bleeding algorithms with high sensitivities which can be used in conjunction with chart reviews to ascertain events within populations of interest.

Hosted file

pds-23-0490-File001.docx available at <https://authorea.com/users/765308/articles/774869-validation-of-a-major-and-clinically-relevant-non-major-bleeding-phenotyping-algorithm-on-electronic-health-records>