Introduction

Chronic urticaria (CU) has been shown to be associated with clinical remission independent of treatment. A machine learning model implementing random survival forest was previously developed to predict time to clinical remission for patients with CU using real-world electronic health records (EHR). The model (PREDICT-CU) contains 176 demographic and clinical variables which is too high for implementation in clinical settings. This study aimed to refine and simplify the model while trying to maintain the accuracy of the prediction.

Materials and Methods

The simplified model, i.e., a partially aggregated model (c-index = 0.62), was developed from the initial PREDICT-CU using the same database (US Optum Life Science EHR [Q1 2007-Q2 2019]) and study cohort of 112,443 patients meeting inclusion criteria. The simplified model’s performance was assessed using concordance as measured by Harrell’s c-index (values reported in the literature 0.6 - 0.75), which represents the degree of agreement between observed and predicted time to remission for all possible pairs of patients. The 20 most important variables from PREDICT-CU were selected and discussed with clinical experts for possible pairs of patients. The simplified model retained 9 clinical and demographic variables (values reported in the literature 0.6 - 0.75), which require specific input and are expected to be available in clinical settings, while medians were imputed for the remaining 167 variables. Following the review of the 20 most important variables, the final simplified model retained 9 clinical and demographic variables which require specific input and are expected to be available in various clinical settings, while medians were imputed for the remaining 167 variables.

Results

The simplified model was developed using the same database (US Optum Life Science EHR [Q1 2007-Q2 2019]) and study cohort of 112,443 patients meeting inclusion criteria. The simplified model retained 9 clinical and demographic variables which require specific input and are expected to be available in various clinical settings, while medians were imputed for the remaining 167 variables. Following the review of the 20 most important variables, the final simplified model retained 9 clinical and demographic variables which require specific input and are expected to be available in various clinical settings, while medians were imputed for the remaining 167 variables. The simplified model may have limited generalizability outside commercially insured US populations and might need further validation in other databases, but it could represent a flexible, portable, and user-friendly tool with wide application across clinical practices.

Discussion

In this study, a complex 176-variable machine learning model was simplified into a 9-variable model that uses clinically available information from patient charts to rapidly predict time to CU remission. The output would be a probability of clinical remission either at prespecified time points (e.g., x% chance to remit by XX months from diagnosis) or within a range of time points (e.g., x% chance to remit between X and Y months). The simplified model may have limited generalizability outside commercially insured US populations and might need further validation in other databases, but it could represent a flexible, portable, and user-friendly tool with wide application across clinical practices.

References


Acknowledgements

The authors acknowledge Priyanka Singh (Novartis, Hyderabad) for editorial and medical writing support. The final responsibility for the content lies with the authors.


To download a copy of this poster, visit the web at: https://bit.ly/3Z3lXav

Copies of this poster obtained through Quick Response (QR) code are for personal use only and may not be reproduced without written permission of the authors.