

# Using machine learning approaches to predict timely clinic attendance and the uptake of HIV/STI testing post clinic reminder messages

Xu X<sup>1,2</sup>, Fairley C<sup>1,2</sup>, Chow E<sup>1,2,3</sup>, Lee D<sup>1</sup>, Aung E<sup>1,2</sup>, Zhang L<sup>1,2,4,5</sup>, Ong J<sup>1,2,6</sup>

<sup>1</sup>Melbourne Sexual Health Centre, Carlton VIC, Australia, <sup>2</sup>Central Clinical School, Faculty of Medicine, Nursing and Health Sciences, Monash University, Carlton VIC, Australia, <sup>3</sup>Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne VIC, Australia, <sup>4</sup>China Australia Joint Research Center for Infectious Diseases, School of Public Health, Xi'an Jiaotong University Health Science Centre, Xi'an Shaanxi, People's Republic of China, <sup>5</sup>Department of Epidemiology and Biostatistics, College of Public Health, Zhengzhou University, Zhengzhou Henan, People's Republic of China, <sup>6</sup>Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London England, United Kingdom

**Background:** Timely and regular testing for HIV and sexually transmitted infections (STI) is important for controlling HIV and STI among men who have sex with men (MSM). We aimed to use machine learning algorithms to predict clinic attendance within 30 days and HIV/STI testing uptake among MSM after receiving a reminder message via short message service (SMS) or email.

**Methods:** Our study included 3,044 clinic consultations among MSM within 12 months after receiving an email or SMS reminder at the Melbourne Sexual Health Centre between April 11, 2019, and April 30, 2020. We used predictors existing routinely collected data in clinical settings such as socio-demographic factors (e.g., age and country of birth), sexual practices, HIV infection status, past STI diagnosis, reasons for attendance (e.g., contact of infection, symptom status), drug use, any sex overseas, participating in sex work, and pre-exposure prophylaxis (PrEP) use. We established multiple machine learning models (e.g., support vector machine, k-nearest neighbour, naïve bayes, random forest, gradient boosting machine, XGBoost, and multi-layer perceptron). Performance metrics were the area under the receiver operating characteristic curve (AUC) and F1 score.

**Results:** About 29.5% [899/3,044] attended within 30 days of receiving the clinic reminder messages, and 84.6% [761/899] had HIV/STI testing. The XGBoost model performed best in predicting clinic attendance within 30 days (mean [SD] AUC=62.8% (3.2%); F1 score= 70.8% (1.2%)). The elastic net regression model performed best in predicting HIV/STI testing within 30 days (AUC= 82.7% (6.3%); F1 score=85.3% (1.8%)).

**Conclusions:** Within 30 days after receiving the clinic reminder messages, the clinic attendance in MSM was relatively low. Machine learning-based approaches can predict the use of sexual health services based on existing routinely collected data in clinical settings. Our predictive models could be used to improve HIV/STI re-testing rates among MSM.

## Disclosure of Interest Statement:

The authors declare no competing interests.