

Development and external validation of a web-based risk prediction tool using machine learning algorithms for an individual's risk of HIV and sexually transmitted infections

P36

Xianglong Xu¹⁻³, Zhen Yu^{2,4}, Zongyuan Ge⁴, Eric P.F. Chow^{1,2,5}, Yining Bao³, Jason J. Ong¹⁻³, Wei Li⁶, Jinrong Wu⁷, Christopher K. Fairley¹⁻³, Lei Zhang^{1-3,8}

¹Melbourne Sexual Health Centre, Carlton VIC, Australia ²Central Clinical School, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne VIC, Australia ³China 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 ⁴Monash e-Research Centre, Faculty of Engineering, Airodoc Research, Nvidia AI Technology Research Centre, Monash University, Melbourne VIC, Australia ⁵Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne VIC, Australia ⁶School of public health, Southeast University, Nanjing Jiangsu, People's Republic of China ⁷Research Centre for Data Analytics and Cognition, La Trobe University, Bundoora VIC, Australia ⁸Department of Epidemiology and Biostatistics, College of Public Health, Zhengzhou University, Zhengzhou Henan, People's Republic of China

Background

HIV and sexually transmitted infections (STI) are major global public health concerns. The World Health Organization (WHO) estimated that over one million curable STIs occur every day amongst people aged 15–49 years worldwide.

Screening of asymptomatic individuals is important for diagnosis, treatment, prevention and control of HIV/STI. Barriers to testing include misjudgment of an individual's HIV/STI risk, limited availability of testing and a high cost of testing.

An easily accessible and user-friendly tool that accurately identifies an individual's risk of HIV/STI could form part of the web-based risk predictions program and play a role in risk prediction and personalised risk management.

This study aimed to develop an HIV/STI risk prediction tool using machine learning algorithms.

Method

We used clinic consultations where individuals who were tested for HIV/STI at the Melbourne Sexual Health Centre between March 2, 2015, to December 31, 2018, as the development dataset (training and testing dataset).

We also used two external validation datasets, including data in 2019 as the external 'validation data 1' and data during January 2020 and January 2021 as the external 'validation data 2'.

We developed 34 machine learning models to assess the risk of acquiring HIV, syphilis, gonorrhoea, and chlamydia.

Finally, we created an online tool to generate an individual's risk of HIV/STI.

Results

Our ML-based risk prediction tool named MySTIRisk performed at an acceptable or excellent level on testing datasets (area under the curve (AUC) for HIV= 0.78; syphilis = 0.84; gonorrhoea = 0.78; chlamydia = 0.70) which had stable performance on both external validation data in 2019 (AUC for HIV= 0.79; syphilis = 0.85; gonorrhoea = 0.81; chlamydia = 0.69), and data in 2020-2021 (AUC for HIV= 0.71; syphilis= 0.84; gonorrhoea = 0.79; chlamydia = 0.69). (Figure 1).

The graphical user interface elements of the tool were summarised in Figure 2. The web application collects individual characteristics, processes the collected characteristics, loads the trained machine learning models, calculates a quantitative HIV/STI risk, and displays the results of the risk and recommendations. A prototype version of the tool is available at www.mystirisk.shinyapps.io/mystirisk.

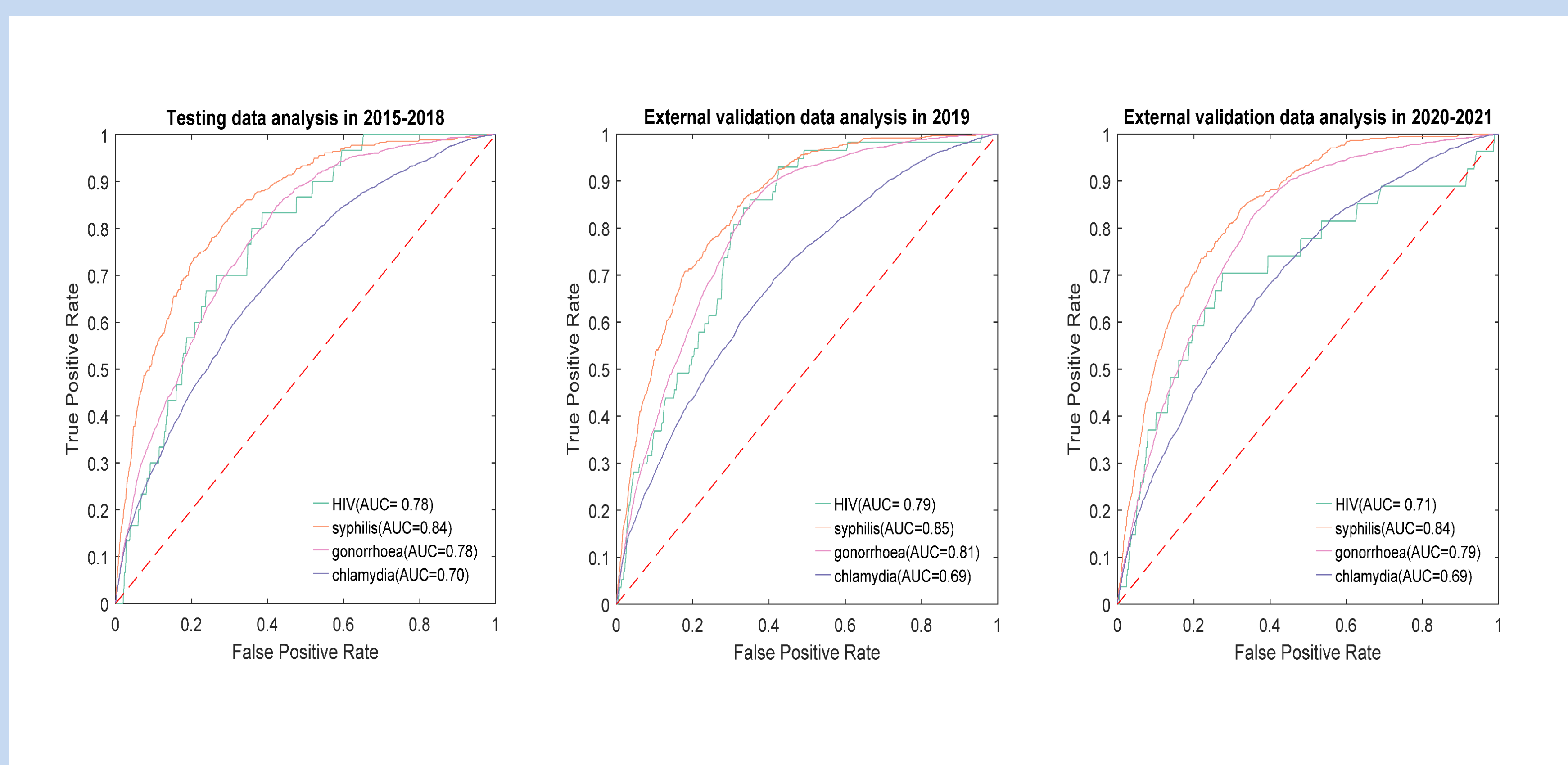


Figure 1. Receiver operating characteristic curve performance of HIV/STI risk prediction tool on testing dataset analysis and external validation analyses. STI: syphilis, gonorrhoea, and chlamydia. AUC: area under the curve.



Figure 2. Graphical user interface elements of the HIV/STI risk prediction tool. STI: syphilis, gonorrhoea, and chlamydia.

Conclusion

This is the first web-based risk prediction tool based on machine learning algorithms and self-reported data to accurately identify HIV, syphilis, gonorrhoea, and chlamydia in men and women and was stable on external validation. Our online risk prediction tool could accurately predict the risk of HIV/STI in clinic attendees with a simple self-administered questionnaire.

MySTIRisk could serve as an HIV/STI screening tool on clinic websites or digital health platforms.

The public can use this tool to assess their HIV/STI risk to inform testing. Clinicians or public health workers can use this tool to identify high-risk individuals for further interventions.

Acknowledgements

We thank Mark Chung at the MSHC for his assistance in preparing figures. Travel expenses of Xianglong Xu were partially covered by the Travel Award sponsored by the open access journal International Journal of Environmental Research and Public Health published by MDPI.

Xianglong Xu also want to thank 2022 Joint Australasian HIV&AIDS and Sexual Health Conference 2022 Joint Australasian HIV&AIDS and Sexual Health Conference Scholarship for supporting the participation in this conference.

Contact

For further communication contact: Xianglong (Eric) Xu
Melbourne Sexual Health Centre
580 Swanston Street, Carlton, VIC 3053, Australia
Email: ericlxu@gmail.com



Joint Australasian
HIV&AIDS +
Sexual Health Conferences
29 August – 1 September

