Prediction of Chlamydia trachomatis retesting and reinfection within one year among heterosexuals with chlamydia attending a sexual health clinic using machine learning approaches

P64

Xianglong Xu[1,2], Eric P.F. Chow [1,2,3], Christopher K. Fairley[1,2], Marcus Chen[1,2], Ivette Aguirre[1], Jane Goller[3], Jane Hocking[3], Natalie Carvalho [4], Lei Zhang[1,2,5,6], Jason J. Ong[1,2,7]

1 Melbourne Sexual Health Centre, The Alfred, Melbourne, Australia 2 Central Clinical School, Monash University, Melbourne, Australia. 3 Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia. 4 Centre for Health Policy, Melbourne School of Population and Global Health, The University of Melbourne. 5 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. 6 Department of Epidemiology and Biostatistics, College of Public Health, Zhengzhou University, Zhengzhou, China. 7 Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

Background

Chlamydia trachomatis is one of the most common bacterial sexually transmitted infections globally, and re-infections are common. Predicting re-testing and re-infection can help direct resources to those who need to be re-tested. Machine learning approaches have some advantages in prediction, including not requiring statistical inferences and assumptions, improving accuracy by exploiting complex interactions between risk factors, handling a mass of predictors and combining them in a nonlinear and highly interactive way. To date, no research has used machine learning algorithms to predict chlamydia re-testing and re-infection. We used machine learning algorithms to predict chlamydia re-testing and re-infection within one year among heterosexuals with chlamydia.

Method

Our baseline data included 6,174 heterosexuals with chlamydia aged ≥18 years old and 2,271 re-tested for chlamydia within one year after their chlamydia diagnosis at the Melbourne Sexual Health Centre from January 2, 2015, to May 15, 2020. We established machine-learning models to predict chlamydia re-testing and re-infection. Chlamydia reinfection was defined as the first new chlamydia diagnosis using nucleic acid amplification testing from any anatomical site, including the oropharynx, urethra/urine, or anorectum, at least 30 days after and within 365 days after a positive chlamydia diagnosis. Machine learning algorithms were conducted with Python 3.9.7. LR, GBM, AdaBoost, GaussianNB, KNN, SVM, RF, and MLP was built using scikit-learn library in Python. XGBoost was built using the xgboost library in Python.

Results

About 36.8% of heterosexual with chlamydia were re-tested for chlamydia within one year, and 15.2% were reinfected with chlamydia. Random forest, gradient boosting machine and XGBoost could predict chlamydia re-testing on the testing data (area under the receiver operating characteristics curve (AUC) > 60.0%). Nonlinear machine learning models performed better than logistic regression in predicting chlamydia re-infection, but they both could not perform well for predicting chlamydia re-infection (AUC < 0.6).

Conclusion

The chlamydia re-testing rate was relatively low within one year among heterosexuals with chlamydia; however, the re-infection rate of chlamydia was high. Our study highlights the need for the innovative interventions to increase chlamydia retesting and reduce reinfection among heterosexuals with chlamydia. To our knowledge, this is the first study using machine learning approaches to predict chlamydia re-testing and re-infection among heterosexuals with chlamydia. Machine learning approaches can improve the prediction of chlamydia re-testing and re-infection compared with traditional logistic regression. We hope our work encourages more machine learning research to explore the effect of introducing additional predictors on predicting chlamydia re-testing and re-infection.

Acknowledgements

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: ericxlxu@gmail.com

Tables and Figures

Table 1. Machine learning model evaluation of chlamydia retesting and reinfection within one year among heterosexuals within one year on the testing data set (mean/SD)

	chlamydia retesting (n=6,174)		chlamydia reinfection (n=2,271)	
	Accuracy, %	AUC , %	Accuracy, %	AUC , %
LR	63.4(3.6)	58.7(0.9)	59.8 (2.5)	50.7 (4.0)
RF	64.1(3.5)	60.8(0.8)	73.0 (1.9)	52.3 (3.2)
KNN	62.0(2.9)	57.5(0.8)	60.2 (3.3)	52.1 (4.4)
Gaussian NB	58.0(0.9)	56.2(0.9)	31.7 (3.1)	49.9 (4.1)
GBM	64.1(4.7)	60.8(1.2)	73.3 (2.0)	51.3 (3.7)
Adaboost	64.2(4.0)	59.7(0.9)	65.6 (2.6)	51.6 (3.4)
SVM	64.0(4.3)	58.3(0.6)	75.7 (2.5)	53.5 (4.0)
XGBoost	64.4(3.9)	60.8(1.5)	73.5 (0.5)	51.9 (1.0)
MLP	63.3(2.9)	58.8(1.1)	73.8 (1.8)	51.6 (4.1)

Note: SD: standard deviation. Logistic Regression (LR), K-Nearest Neighbour (KNN), AdaBoost classifier (AdaBoost), SVM with a Radial Basis Function Kernel (SVM), Gaussian Naive Bayes (GaussianNB), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), Random Forest (RF), and multi-layer perceptron (MLP)

Figure 1. Variable importance in the prediction of chlamydia re-testing within one year among heterosexuals by XGBoost

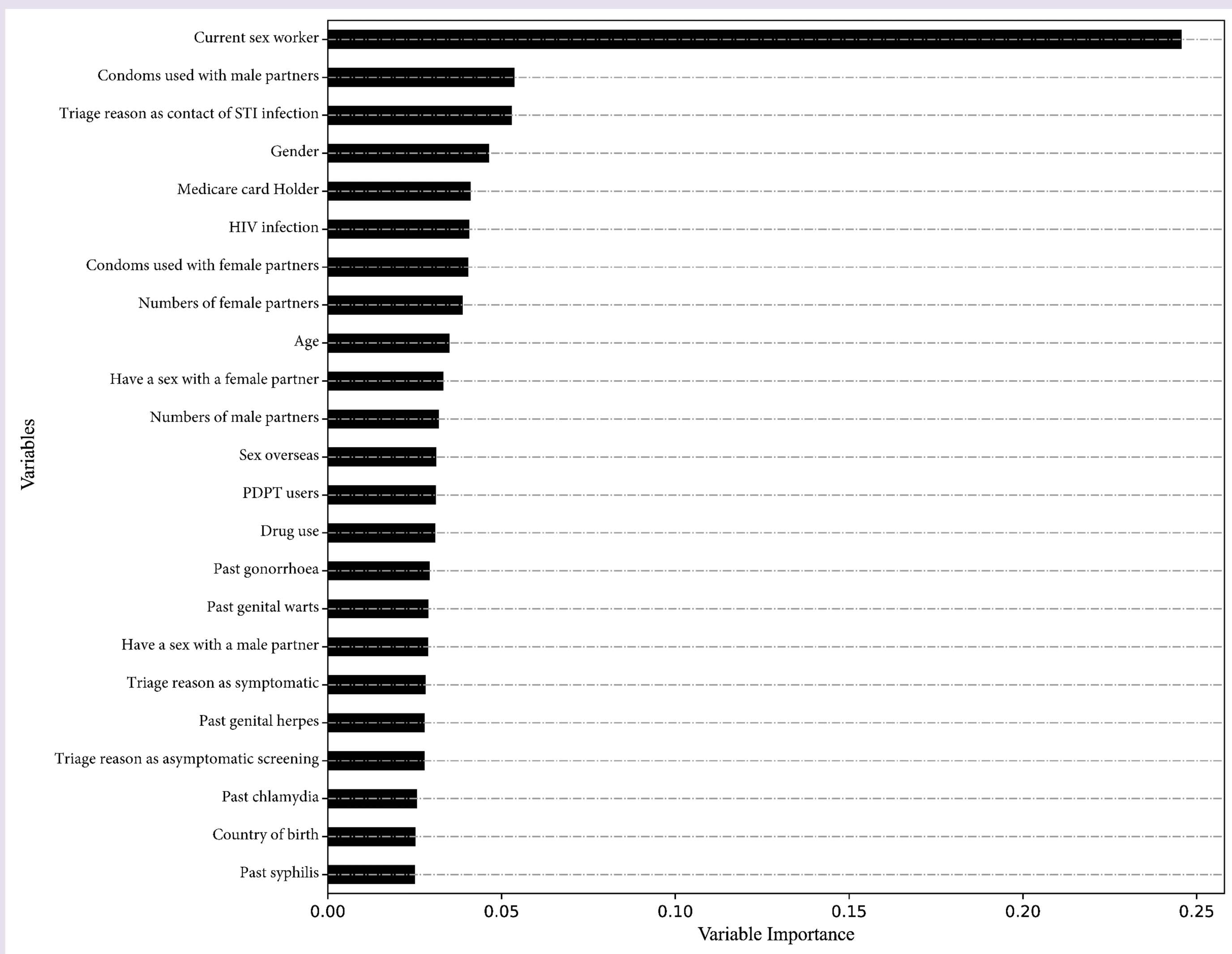

Figure 2. Variable importance in the prediction of chlamydia re-infection within one year among heterosexuals by XGBoost