

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

Xu X<sup>1,2</sup>, Fairley C<sup>1,2</sup>, Chow E<sup>1,2,3</sup>, Chen M<sup>1,2</sup>, Aguirre I<sup>1</sup>, Goller J<sup>3</sup>, Hocking J<sup>3</sup>, Carvalho N<sup>4</sup>, Zhang L<sup>1,2,5,6</sup>, Ong J<sup>1,2,7</sup>

<sup>1</sup>Melbourne Sexual Health Centre, Carlton VIC, Australia, <sup>2</sup>Central Clinical School, Monash University, Melbourne 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>Centre for Health Policy, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne VIC, Australia, <sup>5</sup>Department of Epidemiology and Biostatistics, College of Public Health, Zhengzhou University, Zhengzhou Henan, People's Republic of China, <sup>6</sup>Department of Epidemiology and Biostatistics, College of Public Health, Zhengzhou University, Zhengzhou Henan, People's Republic of China, <sup>7</sup>Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London England, United Kingdom

**Background:** *Chlamydia trachomatis* is the most common bacterial sexually transmitted infection globally, and reinfections are common. Predicting reinfection can help prioritize resources for those who need to be re-tested. Using machine learning approaches, this study aimed to predict chlamydia reinfection within one year among heterosexuals with chlamydia.

**Methods:** Our data included 2271 heterosexuals with chlamydia aged  $\geq 18$  years who 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. 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. Using this data, we established linear and nonlinear machine-learning models to predict chlamydia reinfection. We also used univariable and multivariable logistic regression to conduct risk factor analyses of chlamydia reinfection.

**Results:** In total, 346 (15.2%) heterosexuals with chlamydia were reinfected with chlamydia within one year. Random forest, k-nearest neighbours, gradient boosting machine, adaptive boosting classifier, support vector machine, XGBoost, and multi-layer perceptron performed better than logistic regression in predicting chlamydia reinfection. However, machine learning approaches could not provide a good predictive value for chlamydia reinfection (all area under the receiver operating characteristics curve (AUC)  $< 0.6$ ). Living with HIV (adjusted odds ratios (aOR) = 4.62, 95%CI 2.09-10.09), being a sex worker (aOR= 1.73, 95%CI 1.13-2.63), and older age (aOR= 0.98, 95%CI 0.96-1.00) were associated with reinfection of chlamydia.

## Conclusion:

The reinfection rate of chlamydia is relatively high within one year among heterosexuals with chlamydia. Although machine learning approaches and currently collected data in clinical settings do not offer a good predictive ability for chlamydia reinfection, nonlinear machine learning models have better performance than traditional logistic regression.

## Disclosure of Interest Statement:

The authors declare no competing interests.