EVALUATING A MACHINE LEARNING BASED RISK STRATIFICATION MODEL TO IDENTIFY PEOPLE WHO HAVE INJECTED DRUGS FOR BBV/STI SURVEILLANCE: A RETROSPECTIVE COHORT STUDY

Authors:
El-Hayek C[1,2,3], Nguyen T[1], Asselin J[1], Stoove M[1,2,3], Wilkinson ALW[1,2,3] , Fairley CK[5,6], Read P[4,7], Crawford S[8], Madden A[9], Aung HL[4], Polkinghorne V[1], Boyle DIR[10] , Hocking J[2], Dunn AG[11] and Hellard ME[1,2,3]

[1]Burnet Institute, Melbourne, [2]Melbourne School of Population and Global Health, University of Melbourne, [3]School of Public Health and Preventive Medicine, Monash University; [4]The Kirby Institute, University of New South Wales; [5]Melbourne Sexual Health Centre, Alfred Health, [6]School of Translational Medicine, Monash University, [7]Kirketon Road Centre, South Eastern Sydney Local Health District, [8]Harm Reduction Victoria, [9]International Network of People Who Use Drugs, London, [10]Department of General Practice and Primary Care, University of Melbourne, [11]Biomedical Informatics and Digital Health, Faculty of Medicine and Health, University of Sydney

Background:
Monitoring blood borne viruses (BBVs) and sexually transmissible infections (STIs) among people who have injected drugs is necessary to develop and evaluate healthcare interventions, however injecting drug use (IDU) is not consistently screened for, disclosed or systematically recorded in most healthcare settings. We evaluated a machine learning based risk stratification method that classifies primary care patients based on their likelihood of IDU using a large database of multi-site electronic medical records (EMRs) for surveillance.

Methods:
Using a retrospective sample of de-identified EMRs from the Australian Collaboration of Coordinated Enhanced Sentinel Surveillance of STIs and BBVs, we applied a pre-trained model to generate a probability of IDU for each patient and stratified them as having high- or moderate-probability of having injected drugs. Post-hoc assessment of model behaviour compared the characteristics of model-stratified patients with patients who had pre-existing indications of IDU in their EMRs such as self-report or opioid agonist treatment.

Results:
The study included 1,716,534 patients, of whom 50,235 patients (2.9%) had pre-existing IDU indications. The model classified an additional 118,986 patients (6.9% of total patients), stratified as high- (45,495; 2.7% of total) or moderate- (73,491; 4.3% of total) probability candidates. The main differences between model-classified patient profiles compared to patients with pre-existing IDU indications were a higher proportion of female patients aged 25-39 years, patients visiting general practice clinics specialising in the health of gay and bisexual men, and lower proportion of patients visiting sexual health clinics.

Conclusion:
Our study demonstrated the potential for machine learning to identify patterns within the EMRs of patients likely to have injected drugs. Stratifying and characterising

patients using their probability of IDU may provide an in-depth understanding of patient risk profiles, tailor interventions and improve the reach of healthcare and harm reduction. Further real-world evaluation and refinement of these methods is planned.

Disclosure of Interest Statement:
C El-Hayek has received funding from the National Health and Medical Research Council for this study.