Predicting fatal drug poisoning (overdose) among people living with HIV-HCV co-infection

Mélanie Bédard

Department of Epidemiology, Biostatistics, and Occupation Health, McGill University, Montreal

November 2024

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science in Epidemiology

Table of Contents

ABSTRACT

Background: Drug poisoning (overdose) is an important public health crisis, particularly among people living with HIV and hepatitis C (HIV-HCV) co-infection. Direct-acting antivirals result in high HCV cure rates, successfully reducing liver-related mortality. However, increased rates of drug poisoning deaths will negate these benefits. Investigating the potential predictors for drug poisoning could help to reduce mortality by identifying groups most at-risk.

Objective: The objective of this thesis was to predict six-month drug poisoning mortality among people with HIV-HCV coinfection using socioeconomic, behavioural, and clinical factors, including factors that are routinely measured in clinical practice, as well as those recorded for research purposes.

Methods: Data from the Canadian Co-infection Cohort (CCC) were used. Participants were followed up at six-month intervals when they completed questionnaires on socio-demographic, behavioural and clinical factors. Participants were eligible for analysis if they ever reported injection or non-injection drug use between 2003 and 2023. The outcome was death due to drug poisoning within six months of a participant's cohort visit. We selected a total of 40 predictors. We used a supervised machine learning model, random forest, to develop a classification algorithm. Due to imbalanced data, we used a stratified random forest approach with undersampling. Predictors of drug poisoning were ranked in order of importance and odds ratios (OR) and 95% confidence intervals (CIs) were generated using a generalized estimating equation (GEE) regression with the top five important predictors. Four sensitivity analyses were conducted.

Results: Of 2,132 total CCC participants, 1,998 met the eligibility criteria for this analysis. Of those eligible, 1,764 (88.3%) reported ever using injection drugs and 1,807 (90.4%) reported

ever using non-injection drugs. From a total of 94 drug poisoning deaths, 53 occurred within six months of a participant's last visit. When applied to the out-of-bag sample, the model had an area under the curve (AUC) of 0.61 (95% CI: 0.54, 0.68), indicating poor performance. When applied to the entire sample, the model performed better with an AUC of 0.9965 (95% CI: 0.9941, 0.9988). When ranking the predictors by importance, the top five variables were: addiction therapy in the past six months (6m), history of sexually transmitted infection, smoking (6m), ever being on prescription opioids, and non-injection opioid use (6m). However, the mean decrease in accuracy was low for all variables, indicating that no predictor was very strong. Additionally, the ORs generated by the GEE of the top important variables were close to the null, and almost all 95% CIs associated with these ORs crossed the null, preventing any definitive conclusions to be made on the direction of the association.

Discussion: Ranking variables by importance pointed to some interesting clues as to who might be at risk for fatal drug poisonings, however, due to the challenges we faced in prediction, these results must be interpreted with caution. Our model performed poorly when withholding a sample of the data, and even the most important predictors had little impact on the overall accuracy. These results suggest that drug poisoning deaths may be a random event within the cohort and could reflect the toxicity of the drug supply. Alternatively, the low number of events and imbalanced data would benefit from exploring alternative approaches to investigate this question.

Conclusion: Understanding the predictors of short-term risk of drug poisoning is an important first step for developing clinical tools to target at-risk patients. However, our model performed relatively poorly. As we are unable to identify specific predictors of who is most at risk, efforts need to be placed elsewhere, such as interventions to reduce the toxicity of the supply.

RÉSUMÉ

Contexte: L'intoxication médicamenteuse (surdose) est une crise de santé publique, en particulier chez les personnes avec co-infection de VIH et l'hépatite C (VIH-VHC). Les antiviraux à action directe entraînent des taux élevés de guérison du VHC. Cependant, l'augmentation des taux de décès par intoxication médicamenteuse annulera ces avantages. L'étude des facteurs prédictifs d'intoxication médicamenteuse pourrait contribuer à réduire la mortalité.

Objectif: L'objectif de cette thèse était de prédire la mortalité par intoxication médicamenteuse à six mois chez les personnes avec une co-infection de VIH-VHC en utilisant des facteurs sociodémographique, comportementaux et cliniques.

Méthodes: Les données de la Cohorte canadienne sur la co-infection (CCC) ont été utilisées. Les participants ont été suivis à intervalles de six mois et ont rempli des questionnaires sur des facteurs sociodémographiques, comportementaux et cliniques. Les participants étaient éligibles à l'analyse s'ils avaient consommé des drogues entre 2003 et 2023. Le résultat d'intérêt était un décès dû à une intoxication médicamenteuse dans les six mois suivant la visite de cohorte d'un participant. Nous avons sélectionné un total de 40 prédicteurs. Nous avons utilisé un modèle d'apprentissage automatique supervisé, la forêt aléatoire, pour développer un algorithme de classification. En raison du déséquilibre des données, nous avons utilisé une forêt aléatoire stratifiée avec sous-échantillonnage. Les prédicteurs d'intoxication médicamenteuse ont été classés par ordre d'importance et les rapports de cotes (OR) et les intervalles de confiance (IC) à 95 % ont été générés à l'aide d'une régression par équation d'estimation généralisée (GEE) avec les cinq prédicteurs les plus importants.

Résultats: Sur un total de 2 132 participants au CCC, 1 998 répondaient aux critères d'éligibilité pour cette analyse. Parmi les personnes admissibles, 1 764 (88,3 %) ont consommé des drogues injectables et 1 807 (90,4 %) des drogues non injectables. Sur un total de 94 décès par intoxication médicamenteuse, 53 sont survenus dans les six mois suivant la visite d'un participant. Lorsqu'il est appliqué à l'échantillon hors sac, le modèle présentait une aire sous la courbe (AUC) de 0,61, ce qui indique de mauvaises performances. Lorsqu'il est appliqué à l'ensemble de l'échantillon, le modèle a obtenu de meilleurs résultats avec une AUC de 0,9965. Lors du classement des prédicteurs par importance, les cinq principales variables étaient: la pharmacothérapie, les infections sexuellement transmissibles, le tabagisme, les opioïdes sur ordonnance et la consommation d'opioïdes sans injection. Cependant, la diminution moyenne de l'exactitude était faible pour toutes les variables. De plus, les OR générés par le GEE des variables les plus importantes étaient proches de la valeur nulle, et presque tous les IC à 95 % associés à ces OR croisent la valeur nulle.

Discussion: Le classement des variables par importance a révélé des indices intéressants sur les personnes à risque d'intoxication médicamenteuse. En raison des nombreux défis auxquels nous avons été confrontés lors de la prévision, ces résultats doivent être interprétés avec prudence. Notre modèle a donné de pauvres résultats lors de la rétention d'un échantillon de données, et même les prédicteurs les plus importants ont eu peu d'impact sur la précision globale. Ces résultats suggèrent que les décès par intoxication médicamenteuse pourraient être un événement aléatoire au sein de la cohorte, ou le nombre d'événements et les données déséquilibrées ont posé trop de défis.

Conclusion: Comprendre les prédicteurs du risque d'intoxication médicamenteuse est une première étape pour développer des outils cliniques visant à cibler les patients à risque.

Cependant, notre modèle a peu fonctionné. Il est possible que des efforts doivent être déployés ailleurs, par exemple redoubler les efforts visant à réduire la toxicité de l'approvisionnement.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Dr. Marina Klein and Dr. Erica Moodie for their support and guidance throughout the thesis process. This journey has been a tremendous learning experience, and I am thankful for their faith in me. Their feedback is always thoughtful and impactful, and I appreciate the time and effort they have put into this project alongside me. I would also like to thank my committee member Dr. Joseph Cox for his advice and feedback. I would like to acknowledge the Research Institute of the McGill University Health Centre for their partial financial support in the form of a studentship. In addition, I would like to thank Dr. Marina Klein for her financial support throughout this degree.

I would like to thank and acknowledge the participants of the Canadian Co-infection Cohort (CCC), whose participation allowed me to conduct this research. I would also like to thank the CCC co-investigators, CCC study coordinators, and all staff who make the CCC such an incredible and important cohort.

I would also like to thank the Klein team at the Research Institute of the McGill University Health Centre for being so welcoming and kind. I would like to acknowledge in particular Shouao (Stan) Wang, Alexandra Looky, Isabelle Robichaud, and Chantal Burelle. They are always happy to help and have provided support at various points throughout this process. I would also like to thank Mariam El Sheikh, PhD candidate in the Klein lab. I truly do not think I could have gotten through the past two years without your unwavering support, all your R help, and your sense of humor. Words cannot express how much you have helped me.

Next, I would like to thank the professors, administrative staff, peers, and friends in the Department of Epidemiology, Biostatistics, and Occupational Health (EBOH) at McGill University for strengthening my passion for epidemiology. To my EBOH friends, thank you, I am so happy we were able to go through this process together. I would also like to thank my friends and support systems outside of Montreal who helped me through the lows and who were also there to celebrate the wins.

Finally, I want to thank my family, particularly my parents, for their support throughout my academic journey. I truly could not have gotten through the ups and downs of this process without them. They mean the world to me, and I could not have gotten to this point without them. Thank you so much for everything, I am eternally grateful to you.

CONTRIBUTION OF AUTHORS

All chapters of this thesis were written by Mélanie Bédard and critically revised and reviewed by Marina B. Klein and Erica E.M. Moodie. Marina B. Klein, Erica E.M. Moodie, and Mélanie Bédard contributed to the study conception and design. Statistical analysis was conducted by Mélanie Bédard with guidance from Erica E.M. Moodie. The manuscript was written by Mélanie Bédard and revised by Marina B. Klein, Erica E.M. Moodie, John Gill, Sharon Walmsley, Valérie Martel-Laferrière, Curtis Cooper, Joseph Cox, and circulated to the Canadian Co-infection Cohort co-investigators.

LIST OF TABLES

Chapter 3.2.

LIST OF FIGURES

Chapter 3.2.

LIST OF APPENDICES

Chapter 3.2.

LIST OF ABBREVIATIONS

HIV: Human immunodeficiency virus
AIDS: Acquired immunodeficiency syndrome
ART: Antiretroviral therapy
STBBI: Sexually transmitted blood borne infections
PWID: People who inject drugs
HCV: Hepatitis C virus
DAA: Direct-acting antiviral
CCC: Canadian Co-infection Cohort
gbMSM: Gay, bisexual, and other men-who-have-sex-with-men
PWUD: People who use drugs
POC: Point-of-care
MSM: Men who have sex with men
U=U: Undetectable = Untransmissible
PrEP: Pre-exposure prophylaxis
PEP: Post-exposure prophylaxis
UNAIDS: Joint United Nations Porgramme on HIV/AIDS
SVR: Sustained virological response
OAT: Opioid agonist therapy
LASSO: Lease absolute shrinkage and selection operator
ESLD: End-stage liver disease
CoDe: Coding of Death in HIV
OOB: Out-of-bag
AUC: Area under the curve
PPV: Positive predictive value
GEE: Generalized estimating equation
IQR: Interquartile range
STI: Sexually transmitted infection
OUD: Opioid use disorder
CDC: United States Centers for Disease Control and Prevention
SMOTE: Synthetic minority oversampling technique
AUROC: Area under receiver operating curve

1. INTRODUCTION

1.1. Rationale

There are approximately 39.9 million people living with human immunodeficiency virus (HIV) around the world, and around 63,000 in Canada (1,2). HIV primarily affects the human immune system by weakening the body's defense mechanism, and without adequate treatment can progress into the more advanced acquired immunodeficiency syndrome (AIDS) (3). Although HIV does not have an available cure, effective and safe treatment options exist. Antiretroviral therapy (ART) is the leading treatment option and works by reducing the amount of HIV in the blood (4). ART has successfully helped to decrease rates of AIDS, hospitalization, and mortality among people living with HIV. HIV is transmitted through direct contact with contaminated blood or bodily fluids and therefore belongs to the category of sexually transmitted and blood borne infections (STBBI) (5). This includes sharing drug injection equipment. In Canada, the highest incidence rates of HIV in 2022 were among people who inject drugs (PWID) (6).

Hepatitis C virus (HCV) is a common co-infection among individuals living with HIV. HCV primarily affects the liver, and like HIV is an STBBI because it is transmitted through sexual contact or exposure to infected blood (7). Sharing drug equipment is a particularly high-risk activity for HCV transmission. An HCV infection can be classified as acute or chronic (7). The duration of the infection determines this classification, with infections lasting less than six months considered to be acute, and those lasting longer considered to be chronic. It is estimated that 50 million people are living with chronic HCV infection around the world (8). Advancements in treatment options have transformed HCV into a curable infection. Direct-acting antivirals (DAAs) are the primary treatment option for HCV in Canada and are safe and

effective, curing upwards of 95% of all infections (9). DAAs are also effective among individuals with HIV-HCV co-infection (10,11).

HIV-HCV co-infection occurs in approximately 20-30% of individuals living with HIV in Canada (12). Individuals living with HIV-HCV co-infection often experience an accelerated and worsened progression of HCV infection (13). Fortunately, existing treatments such as ART and DAAs have improved the quality of life and increased the life expectancy of individuals living with HIV-HCV co-infection. Despite these improvements, harms related to drug use in this population, such as drug poisonings, are an ever-growing concern.

Drug poisonings, also known as overdose, are an important public health concern in Canada. Between 2016 and 2023, there were more than 44,000 opioid-related drug poisoning deaths reported across the country (14). Although opioids are the most common substance involved in drug poisonings, stimulant-related drug poisonings are also cause for concern. Drug poisonings among individuals living with HIV and HCV are a public health threat as rates of drug use in this population are high. According to the World Health Organization, over half of the 2.3 million HIV-HCV co-infections around the world occur in PWID (15).

To better understand the drug poisoning epidemic, and to try to reduce the number of deaths, researchers have attempted to predict drug poisonings, often with the use of machine learning (16,17). Machine leaning is a branch of artificial intelligence that allows machines to learn from existing data without explicitly being programmed (18). It has seen success in several fields, including healthcare (19). Machine learning is a powerful tool that could help combat the drug

poisoning epidemic occurring in the country. To date, there are no studies that have attempted to predict drug poisoning deaths in an HIV-HCV co-infected population. In this thesis, I attempt to accomplish this using data from the Canadian co-infection cohort (CCC), one of the largest and longest-running cohorts focused on individuals living with HIV-HCV co-infection (20).

1.2. Research objective

The objective of this thesis is to develop a predictive model for drug poisoning deaths among individuals with HIV-HCV co-infection within six months of a participant's cohort visit using socioeconomic, behavioural, and clinical variables. These variables include a mix of factors routinely measured in clinical practice and those recorded in a research context.

2. LITERATURE REVIEW

2.1. Human immunodeficiency virus

HIV is a virus of the *Retroviridae* family that primarily affects the human immune system, weakening the body's defense mechanism against infections (3). HIV is a lifelong infection that can progress into AIDS, which is more severe, without proper treatment (3). HIV belongs to the category of STBBIs, which are classified by their similar mode of transmission (21). HIV is transmitted through direct contact with infected blood, semen, vaginal fluid, rectal fluid, and human milk (3).

Symptoms of early HIV infection include flu-like symptoms such as chills, fever, fatigue, joint pain, sore throat, headache, and muscle aches (3). These symptoms typically last a few days to a few weeks and can resolve on their own. Many people infected with HIV are asymptomatic and may not be aware that they have been infected until several years later, often when they are experiencing symptoms of AIDS (3). HIV infection results in progressive immune damage increasing the risk of acquiring several other illnesses, also referred to as opportunistic infections and malignancies. Opportunistic infections include tuberculosis, pneumocystis jiroveci pneumonia, and toxoplasmosis (22–24). However, due to effective HIV treatment, AIDS is less common among people living with HIV today.

2.1.1. Risk factors and key populations for HIV

HIV can affect anyone, however there are certain risk factors that increase someone's chances of acquiring the virus. Sexual risk factors such as condomless anal or vaginal sex, a history of another sexually transmitted infection, and harmful use of drugs and alcohol during sexual

activity are important facilitators of HIV transmission (25). Situations where an individual can come into contact with another person's blood, such as sharing drug paraphernalia including needles, syringes, and other injection equipment, and accidental needle injuries can increase one's risk of HIV. Finally, receiving unsafe medical care, such as injections, blood transfusions, or transplants using unsterilized or dirty medical equipment pose an increased risk of infection (25). Due to these risk factors, there are key populations that have a disproportionate burden of HIV infection. These key populations include gay, bisexual, and other men-who-have-sex-with-men (gbMSM), transgender people, sex workers, people who use drugs (PWUD), PWID, and people in the prison system (26).

2.1.2. HIV screening and diagnosis

In Canada, laboratory testing and rapid testing are used to screen for HIV. Both laboratory and rapid testing require blood samples (27). Results from laboratory testing are typically received after one to two weeks, while results from rapid testing are received in minutes. If a laboratory test or rapid test produces a positive test, a second test known as a confirmatory test must be conducted. This confirmatory test will confirm an HIV diagnosis (27).

The two types of rapid testing in Canada are point-of-care (POC) testing and self-testing. POC testing is conducted by a trained professional while self-testing is done by oneself. Rapid tests have the potential to reach those who remain undiagnosed or who are at risk of acquiring HIV (28). A scoping review investigating the use of POC testing in Canada identified a variety of settings where rapid tests were offered such as addiction facilities, dental offices, prisons, commercial sex markets, and on the street through street outreach (29). Early testing is the best

way to ensure diagnosis before progression to AIDS, decrease the risk of transmission, and reduce complications of infection (30). A recent cohort study of men who have sex with men (MSM) in Taiwan found that those who were diagnosed early with HIV experienced a net gain of 8.28 quality-adjusted life years compared to those diagnosed late, in the context of universal access to ART treatment (31).

2.1.3. HIV treatment and prevention

There is currently no cure for HIV, however effective treatments exist. ART drugs stop the virus from replicating, reducing the amount of HIV (also known as the viral load) in the blood to undetectable levels, and slows the spread of the virus in the body (3,25). There is no risk of sexual transmission when people take ART and have an undetectable viral load. Consequently, ART is a cornerstone of HIV prevention. The U=U (Undetectable = Untransmissible) campaign aims to spread this message and reduce the stigma, discrimination, and misinformation around HIV infection (3,32). Many other forms of prevention exist such as condoms and other physical barriers during sex, HIV pre-exposure prophylaxis (PrEP), HIV post-exposure prophylaxis (PEP), safe injecting and smoking supplies, supervised consumption sites, opioid agonist therapy, and safer tattooing and piercing practices (33).

2.1.4. Epidemiology of HIV

According to the Joint United Nations Programme on HIV/AIDS (UNAIDS), it is estimated that 39 million people were living with HIV in 2022 (34). Globally, there were an estimated 1.3 million new HIV infections, and 630,000 AIDS-related deaths in 2022. In Canada, there were approximately 63,000 people living with HIV, and 1,520 new HIV infections in 2020 (35). Of

these new infections in Canada, the majority occurred among key populations, with 43.8% occurring among gbMSM, and 19.8% occurring among PWID (35).

In 2014, UNAIDS committed to ending the AIDS epidemic by 2030, and set the 90-90-90 HIV targets, which were subsequently updated in 2020 to the 95-95-95 HIV targets, in order to achieve this goal (36). The 90-90-90 targets represent the following goals: 90% of people living with HIV knowing their status, 90% of people who know their status accessing treatment, and 90% of people on treatment having a suppressed viral load. In 2022, the Public Health Agency of Canada released a report on their progress towards the 90-90-90 targets, and stated that Canada had achieved both the first and third of the three targets by 2020 (2). As of 2023, several countries have already reached, or are close to reaching the 95-95-95 targets (37,38). As the deadline is approaching, many resources and efforts are still needed to end the epidemic.

2.2. Hepatitis C virus

HCV is a virus in the *Flaviviridae* family that primarily affects the liver (39). HCV is an enveloped virus with a single-stranded RNA genome and measures around 50 nanometers in diameter (40). It is transmitted through direct contact with blood from an infected individual and like HIV, it is considered an STBBI. HCV infection can be described as acute or chronic. Symptoms of acute HCV infection include fatigue, fever, nausea, jaundice, muscle pain, and joint pain, however many people will not have any symptoms (39). Some acute HCV infections will clear spontaneously, however, in about 50%-85% of cases it will progress to chronic HCV without access to proper treatment (39). Chronic HCV can cause liver complications such as

liver disease, cirrhosis, liver failure, or liver cancer. Fortunately, due to advancements in HCV treatment options, namely DAAs, HCV is now a curable infection (41).

2.2.1. HCV risk factors and key populations

The risk factors for HCV are very similar to those of HIV due to shared routes of transmission. Similar to HIV, coming into contact with infected blood can increase the risk of transmission (41). Sharing drug paraphernalia such as needles and syringes, using unsterilized medical equipment, and getting unsafe blood transfusions all pose a risk of HCV transmission (7). HCV can also be transmitted sexually through condomless sex, particularly when blood is involved. HCV has a disproportionately high burden among certain key populations. These populations include PWID, PWUD, people from countries where HCV is endemic, people in the prison system, Indigenous Peoples, and gbMSM (7). Clearance of HCV does not confer immunity from repeat infection; therefore, individuals can become reinfected if they are re-exposed to the virus after cure. Rates of reinfection are highest among individuals who continue to engage in high risk activities, particularly injection drug use (42).

2.2.2. HCV screening & diagnosis

In Canada, individuals are recommended to be screened for HCV using a risk-based approach (7). Risk factors include using shared drug equipment, exposure to non-sterile medical equipment, sexual activity where blood may be involved, living or visiting a country with high HCV prevalence, and being born to a person with HCV. All currently approved HCV tests in Canada require a blood sample. A first test detects the presence of antibodies to HCV indicating previous exposure to the virus (41). A second test detects the virus in the blood, typically with an

RNA test or a core antigen test, which is required to confirm chronic infection (41). If a screening test is negative, no further testing is required, unless there are ongoing risk factors. After HCV cure, HCV RNA test will become negative however, the antibody test remains positive lifelong. Detecting reinfections requires repeated HCV RNA testing.

2.2.3. HCV treatment & prevention

DAAs are the current treatment for HCV infection in Canada (41). In order for an individual with HCV to be cured, they must achieve sustained virological response (SVR), where a negative or undetectable HCV RNA test is obtained. DAAs are highly effective and can cure upwards of 95% of people living with HCV infection. In addition to the high cure rates, DAAs are also beneficial in reducing liver-related complications, such as fibrosis progression, portal hypertension, and end-stage liver disease (ESLD), as well as several non liver-related complications (43,44). DAAs have also reduced mortality among people living with HCV, increasing life expectancy in this population (45). In an American cohort study following individuals seeking HCV care, a significant decrease in mortality was observed among individuals who had previously completed DAA treatment compared to individuals who did not receive DAA treatment (45).

Harm reduction is an important form of prevention for HCV infection, particularly among the key populations of PWID and PWUD (41). This includes, but is not limited to, access to clean drug equipment, education, and supportive environments. Condoms and other physical barriers during sex, safer tattoo and piercing practices, and avoiding sharing drug equipment are other recommended prevention practices (41).

2.2.4. Epidemiology of HCV

It is estimated that 50 million people worldwide are living with chronic HCV infection (8). There are approximately 1 million new HCV infections globally every year. In 2022, it was estimated that 242,000 people died from chronic HCV infection (8). In Canada, it is estimated that 204,000 people were living with chronic HCV infection in 2019 (46). There were 9,470 new cases of HCV infection in Canada in the same year. In 2016, the World Health Assembly of the WHO set the goal of eliminating viral hepatitis as a major public health threat by 2030 (47). To achieve this goal, the targets of reducing new chronic viral hepatitis infections and deaths by 90% and 65% respectively between 2016 and 2030 were established (47).

2.3. HIV-HCV co-infection

People living with HIV are at an increased risk of HCV infection. HIV and HCV have similar modes of transmission and key populations. HIV-HCV co-infection is greatest among PWID (48). It is estimated that there are over 2 million people worldwide living with HIV-HCV co-infection, and approximately 14,000 in Canada (48,49). A meta-analysis conducted by Platt et al. found that people living with HIV are six times more likely to be infected with HCV than their HIV-negative counterparts (48). Additionally, HCV disease progression occurs faster among those with HIV compared to those without. Higher HCV viral load, increased risk of chronic HCV infection, more rapid progression to liver fibrosis and cirrhosis, are features of co-infection (13). Fortunately, DAAs work equally as well for those with HIV-HCV co-infection as those with HCV monoinfection (10,11,50). Additionally, DAAs have reduced both liver and non-liver related outcomes in this population. An Italian retrospective observational study observed high efficacy and safety rates of DAAs among the co-infected population, while a retrospective

analysis of clinical data in Spain found that co-infected individuals treated with DAAs experienced similar improvements in survival and the onset of comorbidities compared to individuals with HCV monoinfection treated with DAAs (11,51).

Advancements in effective treatment options for both HIV and HCV have led to gains in life expectancy among these populations. The introduction of ART has turned HIV into a manageable chronic condition, while DAAs have provided a cure for HCV. Prior to these advancements, liver-related mortality was the most important cause of death among the HIV-HCV co-infected population (52). However, drug poisonings, also known as overdose, have become a rapidly growing cause of death among individuals living with HIV-HCV. In the Canadian Co-infection Cohort, a multicenter prospective cohort study following individuals with HIV-HCV co-infection across Canada, drug poisoning was identified as the most common cause of death overall (52). Additionally, a retrospective cohort conducted in British Columbia found that the drug poisoning epidemic decreased the life expectancy of people living with HIV, reducing the life expectancy of a 20 year old with HIV by over three years between 2014 and 2017 (53). Drug poisoning deaths threaten to overshadow the gains made in life expectancy due to HIV and HCV treatment in this population.

2.4. Drug poisoning (overdose)

Drug poisoning, also known as overdose, occurs when an individual “takes one or more drugs in a quantity or combination that exceeds what their body can handle” (54). Drug poisonings caused by central nervous system depressants (such as opioids) are often characterized by the following symptoms: unresponsiveness, shallow or irregular breathing, slowed heart rate,

seizures, and vomiting. Drug poisonings caused by central nervous system stimulants (such as amphetamines) are characterized by the following: tremors, flushed skin, headaches, panic, paranoia, confusion (54). In the most severe cases, drug poisonings can result in death; these are considered fatal drug poisonings.

In recent years, the rate of fatal drug poisonings in Canada has been increasing at an alarming rate. In 2016, British Columbia declared a public health emergency in response to the massive increase in drug poisonings and drug poisoning deaths (55). April 2024 marked the eighth anniversary of this declaration, and the emergency is still ever-present (56). In a statement released by the Public Health Agency of Canada, opioid-related harms in the first three quarters of 2023 were similar to those in the peak of the COVID-19 pandemic in 2021, the highest recorded since the start of collection in 2016 (57). There have been 44,592 opioid poisoning deaths across the country between January 2016 and December 2023 (14). Although opioids are responsible for most fatal drug poisoning events, psychostimulants also contribute an important portion of these events (14).

2.4.1. Risk factors associated with drug poisonings

There are several factors that can increase someone's risk of experiencing a drug poisoning event. A change in tolerance due to reduced or no drug consumption can lead someone to use a higher dose than they are used to (58). This can occur after an individual has been in jail for an extended period or has recently completed an addiction therapy program (59,60). A change in the drug supply can lead an individual to use a drug that has been cut with another drug that they had not initially intended to take. Evidence suggests that border closures during the peak of the

COVID-19 pandemic affected the availability of drugs, which caused an increase in drug alterations (61). Altering drugs by adding foreign substances and other drugs pose a serious risk to the health of PWUD. This is known as unintentional polysubstance use. The intentional mixing of drugs, known as intentional polysubstance use, is also a risk factor for drug poisonings (62). Central nervous system depressants, known as "downers" are often mixed with central nervous system stimulants, known as "uppers" which is an especially fatal mix. Mixing these two types of drugs can modify, and often mask, the effect of the drugs, making it easier to experience a drug poisoning event (63). A history of past drug poisoning events is also a risk factor associated with a fatal drug poisoning event. A study investigating risk factors for fatal drug poisonings using data from two prospective cohorts in Vancouver, British-Columbia found that experiencing a non-fatal drug poisoning posed an elevated risk of experiencing a fatal drug poisoning (64). The authors also found that the greater the number of non-fatal drug poisoning events, the greater the risk of fatal drug poisoning. Additional factors that have been identified in the literature to be associated with drug poisonings include homelessness, previous experience in the prison system, low income, gender, age, race, unmet mental health needs, polysubstance drug use, history of opioid prescription, and opioid use disorder treatment discontinuation (65–68).

2.4.2. Drug poisonings, HIV, and HCV

Due to the large number of individuals living with HIV-HCV co-infection who are also PWUD, drug poisonings are an important public health threat to this population. Several studies have investigated drug poisonings among people living with HIV and people living with HCV. Current evidence suggests that people living with HIV are at a greater risk of experiencing a drug poisoning event (59,69). A systematic review and meta-analysis conducted by Green et al. found

a positive association between positive HIV status and drug poisoning (59). The literature suggests that this relationship may be explained by the high rate of chronic pain among individuals living with HIV, leading to an increased likelihood of receiving prescribed opioids (69). Additionally, the risk of both HIV and overdose are highest among those who face discrimination at a social and structural level (70).

The literature also suggests that people with HCV are more likely to experience a drug poisoning event than those not infected with HCV (71,72). In a cohort in British Columbia investigating drug-related deaths among people living with HCV and their HCV-negative counterparts, drug-related deaths were higher among those with HCV (71). The drug poisoning mortality trends among people with HCV also appear to reinforce this (72). A study investigating the trends in mortality for HCV and alcoholic liver disease in the United States between 2009 and 2018 found the age-standardized drug overdose mortality increased at an annual rate of 3% among people with HCV (72). A cross-sectional study in Tennessee found that the prevalence of HCV was 24.5% among drug poisoning deaths from all drugs and 35.4% among methamphetamine and opioid-related drug poisoning deaths (73). This evidence highlights the drug poisoning crisis among people living with HCV.

Authors Perlman & Jordan describe HIV, HCV, opioid misuse and overdose as a syndemic (74). In short, a syndemic is defined as two or more co-occurring diseases/health conditions in a population that interact and negatively exacerbate the other due to underlying biological, social, economic, or environmental conditions (74). The relationship between HIV, HCV, and drug

poisoning is very complex and multilayered. However, it is evident that it is an important public health issue that requires many resources to combat.

2.4.3. Harm reduction

Harm reduction is an important initiative to reduce drug poisonings and other harms caused by drug use. At its core, the goal of harm reduction is not necessarily for PWUD to stop or reduce their substance use, but instead to minimize the health and/or social harms related to substance use and addiction (75). In Canada, harm reduction includes, but is not limited to, safe consumptions sites, overdose prevention services, naloxone distribution, and drug checking (76). Harm reduction also helps to reduce both HIV and HCV transmission. Evidence compiled in a review by Palmateer and colleagues in 2022 suggests that opioid agonist therapy (OAT) reduces HIV and HCV transmission, while needle and syringe programs reduce HIV transmission, and a combination of OAT and needle and syringe programs reduces HCV transmission (77). There are many benefits of harm reduction that can be leveraged for PWUD in the context of drug poisonings and STBBIs such as HIV and HCV. With reduced use of harm reduction services, the risk of HCV infection, HIV infection, and drug poisoning is greater.

2.5. Machine learning

Machine learning, under the larger umbrella of artificial intelligence, is a branch of computer science that has gained popularity in recent years in part due to the increased availability and accessibility of data (78). Machine learning is defined broadly as giving computers the ability to learn and function intelligently without being explicitly programed (79). It is used in many industries and sectors such as cybersecurity, agriculture, finance, and healthcare (79). Machine

learning relies on data to learn the patterns of individuals, businesses, and events, thus the better the data, the better the model will perform (79,80). Machine learning algorithms range from simple and straightforward to very complex. One popular machine learning algorithm is random forest, which combines the output of several classification trees to inform one final result (81). Classification trees, also known as decision trees, are used to predict an outcome using a series of binary decisions that split data into smaller and smaller partitions based on various input variables (78). By combining the output of several classification trees, the results of the random forest are typically considered to be more accurate and generalizable compared to a single classification tree (81).

2.5.1. Machine learning and artificial intelligence in a healthcare context

Artificial intelligence and machine learning have become increasingly popular in healthcare settings and they have shown to be helpful in several aspects of the field such as medical imaging, diagnosing, and disease prediction (82). A systematic review found that among nine studies comparing artificial intelligence to medical experts in disease diagnosis, artificial intelligence had comparable performance to the human experts (83). Healthcare data and their sources vary widely, offering ample opportunities to successfully incorporate machine learning into everyday clinical practice (84). Data from electronic medical records, genomics, and medical imaging have been utilized to develop machine learning algorithms (84). For example, an American study used electronic health records and emergency department triage data to develop an machine learning algorithm to predict hospital admission at the time of emergency department triage (85).

Although there have been many successes when incorporating machine learning in a healthcare context, there have also been several hurdles. For instance, artificial intelligence and machine learning algorithms used for identifying skin diseases have been criticized for underperforming on darker skin tones (86). Researchers warn that ethical concerns, transparency, and interpretability must be considered when generating machine learning algorithms for a healthcare context (87).

2.5.2. Predicting drug poisoning

The overwhelming number of deaths due to drug poisoning have prompted several researchers to attempt to predict these events in an effort to prevent future deaths. The narrative review by Tseregounis & Henry identified articles that generated clinical prediction models for opioid drug poisonings (88). The c-statistic, a measure of the goodness of fit, of the 12 included studies ranged from 0.69 (fair) to 0.95 (excellent). Existing literature on predicting drug poisoning has used a variety of methods such as random forest, least absolute shrinkage and selection operator (LASSO), logistic regression, and cox proportional-hazards model (88). Almost all included studies reported high negative predictive values and low positive predictive values, resulting in a high number of false positives. This was due to drug poisoning being a rare event amongst the datasets used (88), which can create distinct challenges in accurate classification of the minority class. In particular, when the minority class (drug poisonings) are very rare, a model can appear "highly accurate" even if all predicted outcomes were made as "not drug poisoning" since only a small fraction of the total outcomes would be misclassified. These issues are discussed in greater detail in the methods section of the manuscript.

2.6. Summary

In this chapter, I reviewed the epidemiology of HIV and HCV, and discussed the risk factors, key populations, screening methods, treatment options, and prevention strategies associated with these infections. I also touched on the details of HIV-HCV co-infection. I reviewed the current state of drug poisonings in Canada, harm reduction strategies, and drug poisonings among people living with HIV-HCV co-infection. Finally, I explored the topic of machine learning, including the benefits of machine learning in healthcare settings, and leveraging it to predict drug poisoning events. This chapter touched briefly on challenges that have been observed in the literature studying prediction of drug poisonings; many core issues of this topic will be further addressed in the methods section of chapter 3, which contains details of the dataset and analytic methods used in the manuscript.

3. PREDICTING FATAL DRUG POISONING AMONG PEOPLE LIVING WITH HIV-HCV CO-INFECTION

3.1. Preamble

Drug poisoning is a serious public health concern that is affecting the lives of many Canadians, including individuals who are living with HIV-HCV co-infection (14,52). The use of machine learning algorithms, such as random forest, to attempt to predict fatal drug poisoning events could help to combat the high death rates and help to uncover the characteristics of individuals who are especially at-risk. A previous narrative review identified studies that have developed predictive models to predict drug poisoning events with a range of success, however, no existing studies focus on predicting these events within the HIV-HCV co-infected population (88). In this manuscript, I developed a random forest classification algorithm using data from the Canadian Co-infection Cohort to predict fatal drug poisonings among individuals with HIV-HCV co-infection. I also conducted several sensitivity analyses to attempt to predict drug poisoning events in different contexts. The results of this thesis are presented in one manuscript. This manuscript is to be submitted for publication in the Canadian Liver Journal.

3.2. Manuscript

Title page

Title: Predicting fatal drug poisoning among people living with HIV-HCV co-infection

Short running title (55 characters or less): Drug poisoning and HIV-HCV co-infection

Authors: Mélanie Bédard[1,2], Erica E.M. Moodie[1], Joseph Cox[1,2], John Gill[3], Sharon Walmsley[4,5], Valérie Martel-Laferrière[6], Curtis Cooper[7,8], Marina B. Klein[1,2], and the Canadian Co-infection Cohort Investigators.

Affiliations:

1. Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada
2. Centre for Outcomes Research and Evaluation, McGill University Health Center-Research Institute, Montreal, QC, Canada
3. Department of Medicine, University of Calgary, Calgary, AB, Canada
4. Division of Infectious Diseases, Department of Medicine, University of Toronto, Toronto, ON, Canada.
5. Toronto General Hospital Research Institute, University Health Network, Toronto, ON, Canada
6. Centre de recherche du Centre Hospitalier de l'Université de Montréal, Montreal, QC, Canada
7. Department of Medicine, University of Ottawa, Ottawa, ON, Canada.
8. Ottawa Hospital Research Institute, Ottawa, ON, Canada.

Institution from which the work originated: McGill University and McGill University Health Centre-Research Institute

Corresponding author:

Marina B. Klein

McGill University Health Centre

Division of Infectious Diseases and Chronic Viral Illness Service

5252 Blvd. de Maisonneuve

Montreal, QC, Canada H4A 3S9

Email: marina.klein@mcgill.ca

Structured abstract (250 words)

Background: Drug poisoning (overdose) is a public health crisis, particularly among people living with HIV and hepatitis C (HCV) co-infection. Identifying potential predictors of drug poisoning could help decrease drug-related deaths.

Methods: Data from the Canadian Co-infection Cohort were used to predict death due to drug poisoning within six months (6m) of a cohort visit. Participants were eligible for analysis if they ever reported drug use. Supervised machine learning (stratified random forest with under-sampling to account for imbalanced data) was used to develop a classification algorithm using 40 sociodemographic, behavioural and clinical variables. Predictors were ranked in order of importance and odds ratios and 95% confidence intervals (CIs) were generated using a generalized estimating equation regression.

Results: Of 2,175 study participants, 1,998 met the eligibility criteria. There were 94 drug poisoning deaths, 53 within 6m of a last visit. When applied to the entire sample, the model had an area under the curve (AUC) of 0.9965 (95% CI, 0.9941, 0.9988). However, the false positive rate was high resulting in a poor positive predictive value (1.5%). Our model did not generalize well out of sample (AUC 0.6, 95% CI 0.54, 0.68). The top important variables were addiction therapy (6m), history of sexually transmitted infection, smoking (6m), ever being on prescription opioids, and non-injection opioid use (6m). However, no predictor was strong.

Conclusions: Despite rich data, our model was not able to accurately predict drug poisoning deaths. Larger datasets and information about changing drug markets could help improve future prediction efforts.

Keywords: HIV-HCV co-infection, drug poisoning, overdose, random forest, machine learning

Lay summary (Recommended length around 300 words)

Drug poisoning, also known as overdose, is a public health threat that affects the lives of many Canadians. Drug poisonings have been on the rise since 2016, with a notable increase observed following the COVID-19 pandemic. Drug poisoning deaths have also been increasing among people who are co-infected with HIV and hepatitis C (HCV). In an attempt to identify those at higher risk of harm from drug poisoning, the goal of this analysis was to predict drug poisoning deaths within an HIV-HCV co-infected population. Data from the Canadian Co-infection Cohort were used, a study that follows over 2,000 people living with HIV-HCV co-infection. A machine learning method, known as random forest, was used to predict drug poisoning deaths using information collected on participants' sociodemographic, clinical, and behavioural factors. While our model appeared to perform well when run on the entire data set, it was not able to accurately to predict drug poisoning deaths on a subset of data and its predictive value was poor. Potential predictors were ranked by their importance, and the top five important variables were: addiction therapy in the past six months, history of sexually transmitted infection, smoking in the past six months, ever being on prescription opioids, and non-injection opioid use in the past six months. However, none of these variables were very strong predictors. Although our study provides clues as to factors that might predict drug poisonings among people with HIV-HCV co-infection, accurately predicting these events was difficult. Working with larger datasets and having access to detailed information about changing drug markets could help improve future prediction efforts.

Introduction

Hepatitis C virus (HCV) infection is a common co-infection among people living with HIV due to shared routes of transmission (1). People co-infected with HIV-HCV experience more rapid HCV disease progression than those with HCV monoinfection, resulting in increased risks of liver fibrosis, cirrhosis, liver cancer, and death (2). Advances in treatment options have transformed HCV into a curable disease. Direct-acting antivirals (DAAs), introduced in the early 2010s, are now widely available in Canada and have cure rates of over 95% (3). A population-based study found that increased access to DAAs in 2014 led to a decrease in overall HCV-related hospitalizations and an increase in life expectancy among those living with HCV in Canada (3).

Although DAAs have successfully reduced liver-related mortality, deaths due to drug poisonings, also known as overdose, pose an increasing threat to people with HIV-HCV co-infection. A large majority of those living with HIV-HCV co-infection are people who inject drugs (PWID) (1). According to the World Health Organization, over half of the 2.3 million HIV-HCV co-infections around the world occur in PWID (4). In the Canadian Co-infection Cohort (CCC), an open prospective cohort study following individuals living with HIV-HCV co-infection, drug poisoning surpassed end-stage liver disease (ESLD) as the most common recorded cause of death during the period of 2013-2017 (5).

Drug poisoning is an important public health issue in Canada. A notable increase in drug poisoning deaths began in 2016, rising substantially with the onset of the COVID-19 pandemic,

and have remained high ever since (6,7). Between January 2016 and December 2023 there were 44,592 apparent opioid-related drug poisoning deaths across Canada (8).

In recent years, artificial intelligence and machine learning methods have been used in healthcare settings to assist with diagnosis, medical imaging, and disease prediction (9). These technologies can help to speed up processes and reduce the number of errors compared to humans. Predictive analytics, the concept of predicting future outcomes based on previously collected data, is an especially useful aspect of machine learning and has been used in clinical trials and healthcare operations (10). To reduce the high number of drug poisoning deaths, researchers have turned to machine learning to predict these events, with the aim of ultimately preventing deaths. A recent review identified 12 studies that developed clinical prediction models to predict opioid drug poisoning events, with a range of model performance from fair to excellent (11). However, none of the included studies focused on HIV-HCV co-infected populations or substances other than opioids (e.g. stimulants such as methamphetamine and sedatives such as xylazine) which are increasingly implicated in drug poisoning deaths (8).

The objective of this study was to use machine learning to predict drug poisoning mortality among people with HIV-HCV co-infection. We used a large array of socioeconomic, behavioural, and clinical factors, including variables that are routinely measured in clinical practice, as well as those recorded for research purposes.

Methods

Data source

The CCC is an open prospective cohort study following participants living with HIV-HCV co-infection from 18 sites across six provinces since 2003 as described previously (12). Briefly, participants must be ≥ 16 years old, have a documented HIV infection, and have either chronic HCV infection or evidence of HCV exposure. Participants are followed at six-month intervals when they complete questionnaires on socio-demographic, behavioural, substance use and clinical factors. Clinical data are collected from medical chart reviews and blood tests are performed at every visit.

Study population

Participants from the CCC were eligible for this analysis if they ever reported injection drug use or non-injection drug use, excluding cannabis, between April 2003 and July 2023. Drug use was assessed at the baseline visit and at every six-month follow-up visit. Thus, participants could become eligible for analysis if their first report of drug use occurred after baseline. Only visits occurring on and after the first report of drug use were included in the analysis.

Outcome

The primary outcome of interest was death due to drug poisoning (overdose) within six months of a participant's last visit. Since participants are followed at six-month intervals, similar to standard clinical care, the rationale behind predicting the outcome within this interval was that, if successful, a healthcare worker would be able to intervene at any given visit if the risk at that visit was high. In the context of this work, we used the Canadian Mental Health Association's

definition of overdose: “taking one or more drugs in a quantity or combination that exceeds what their body can handle” (13). Causes of death were categorized using the Coding of Death in HIV (CoDe) system (14). Information on deaths was collected in a standardized form reviewed by two independent reviewers. These forms include clinical information as well as information from autopsy and coroner’s reports, when available.

Predictors

We used socio-demographic, behavioural, and clinical variables collected in the CCC as predictors. We initially selected 93 potential predictors based on the literature and expert opinion. However, after conducting cross-tabulations of each potential predictor and the outcome, several of the variables had fewer than five participants in at least one of the cells, such that any analyses including these variables would exhibit unstable estimates. When possible and substantively meaningful, variables were collapsed based on similar categories (i.e., combining drugs considered stimulants). However, this was not always possible; in such cases, variables where data were too sparse were omitted from the list of potential predictors. Following this process of removing variables with sparseness, there remained 40 potential predictors for consideration in our predictive model; these are detailed in Table 1.

Table 1. – Candidate predictors used in the predictive model (random forest classifier)

Category	Predictors
Socio-demographic	Sex, transgender identity, education, race/ethnicity, age, sexual orientation, employment status, income
Clinical	End stage liver disease, on antiretroviral therapy, low CD4 cell count (<200 cells/µl), detectable HIV RNA, detectable HCV RNA, sexually transmitted infection (ever), on prescription benzodiazepines (ever), on prescription opioids (ever)

Behavioural	Alcohol use (6m), hazardous alcohol use (6m), incarceration (6m/ever), sex work (ever), cigarette smoking (6m), cannabis use (6m), age at first injection drug use, injection drug use (6m/ever), sharing injection drug equipment (6m), use of clean needle exchange program (6m), being in therapy or program for addiction (6m), injection opioid use (6m/ever), injection stimulant use (6m), injection polysubstance use (6m), non-injection drug use (6m/ever), age at first non-injection drug use, non-injection opioid use (6m/ever), non-injection polysubstance use (6m), non-injection stimulant use (6m)

*6m = in the past six months

Predictive model: Random forest

We used a random forest model classifier to predict drug poisoning events. Random forest is an ensemble machine learning technique that combines the outputs of multiple classification trees to inform one final outcome (15). Classification trees use information from various input variables to predict an outcome using a series of binary choices (16). In a classification tree, all data starts at the top of the tree in what is called the root node, and the decision rules split the data based on a specific feature, creating smaller and smaller nodes until a stopping criterion is reached. Relying on the results of a single classification tree may be inaccurate in the sense of having poor generalizability, as a single tree is prone to overfitting; taking the majority vote of several classification trees (hence the term "forest") typically yields more reliable results in the sense of better out-of-sample prediction (15). A key feature of the random forest method is bootstrap aggregation, also known as *bagging*. The main idea of bootstrap aggregation is that the classification trees are fit ("trained") using a subsample of the data that is resampled with replacement. In other words, certain samples will appear more often than others in individual classification trees. This helps to avoid overfitting, as the classification trees are not being fit on the entire sample (15). Additionally, random forests also use *feature bagging*, an approach in

which not every input variable is considered at each split in the classification tree, but rather a random subset. Feature bagging has been shown to result in more diversity and less correlation between trees (15). The random forest was chosen for this analysis because it is a powerful machine learning algorithm that is accurate, relatively robust to outliers and noise, and not prone to overfitting (15,17).

There are three key parameters in the random forest that can be tuned to maximize the model's performance: the number of variables randomly sampled at each node split, the number of classification trees in the random forest, and the minimum size of the terminal node (i.e., the smallest number of observations allowed in a node).

Handling imbalanced data

Imbalance in data is a common challenge for random forests using classification algorithms because the standard implementation of random forest prioritizes the prediction accuracy of the majority class (18). Imbalance can be severe with rare events such as deaths, and this was indeed a challenge in our analysis: in the CCC, only 0.3% of events were drug poisoning deaths. Due to the rareness of the outcome, it is very likely that any given bootstrap resample will contain very few or no drug poisoning events. One approach that has been proposed to address outcome class imbalance in random forest is to use stratified sampling with under sampling of the majority class (in this case, no drug poisoning) to improve model performance. Stratified sampling is a technique that partitions the population into strata based on a certain characteristic (typically the outcome class), and random sampling for the bootstrap is then conducted in each strata (19). We

opted to under sample the majority class such that there were an equal number of drug poisonings and non-drug poisonings in each bootstrap iteration.

All analyses were conducted using R version 4.4.0, RStudio 2023.12.1+402 and R packages Caret and randomForest to develop and evaluate the algorithm (20–22). Missing data were imputed using the rfImpute function from the randomForest package in R (20).

Model tuning

The random forest model was tuned using 10-fold cross-validation to select model parameters. The model was trained and tested on the entire dataset; data were not split into separate training and testing datasets to avoid further reducing the already low event rate, and instead we relied on out-of-bag (OOB) error estimates. OOB error estimates and 10-fold cross-validation were used to tune the hyperparameters of the model. We used the accuracy metric to tune the three hyperparameters: the number of variables at each split, the number of trees to fit, and the minimum size of the terminal node. We used the following ranges for the candidate values of the tuning parameters, where p represents the number of candidate predictors: the number of variables at each split: sqrt(p) to p/2, the number of trees: 100 to 1500, the minimum size of the terminal node: 1 to 10. In the case where the lower bound of the number of variables at each split produced the best accuracy, we reduced the range to 3 to see if the accuracy was improved. We selected the candidate values of the tuning parameters that generated the highest accuracy. In the case where there was little or no difference in the accuracy, we opted for the recommended defaults. The recommended defaults for a classification random forest are a number of variables

at each split equal to the square root of the number of candidate predictors, a minimum size of the terminal node of 1, and the number of trees of 500 (20).

Results reported

Results from the random forest classifier are reported in terms of model accuracy, area under the curve (AUC), positive predictive value (PPV), and sensitivity of the final model. We generated variable importance plots, and we conducted a generalized estimating equation (GEE) regression of the top five important variables to provide an interpretable description of the direction and strength of the association between drug poisonings and those identified as important predictors by the classifier. We chose to use a GEE regression with a logistic link and an exchangeable correlation structure to account for the correlation between visits of the same individual.

Sensitivity analysis

Four sensitivity analyses were conducted. First, all drug poisoning deaths up to one year after a participant's last visit were included, allowing for a greater number of drug poisoning outcomes. Second, to take into account the changing drug landscape caused by the opioid epidemic, only visits that occurred in 2016 or later (8) were included in the analytic set. Since this reduced the number of drug poisoning events, certain variables now had fewer than five participants in at least one of the cells, and these additional candidate predictors were eliminated for this analysis, leaving a total of 34 potential predictors. Third, only active drug users were included, defined as participants who had indicated using drugs within the last six months at any visit. Finally, any deaths by an unknown cause within six months of a participant's last visit were reclassified as a drug poisoning death. Although this allowed us to consider more predictors, we focused only on

the 40 predictors identified in the primary analysis. Unknown deaths make up the largest proportion of deaths in the CCC and given that drug poisonings often are not well documented or investigated, it is plausible that many may be due to drug poisoning.

Results

Study population

Of 2,175 study participants, 1,998 met the eligibility criteria and were included in the analysis. The 1,998 participants contributed a total of 17,837 visits. There were 94 drug poisoning deaths, of which 53 occurred within six months of a participant's last visit. Of the eligible participants, 1,764 (88%) reported ever using injection drugs, and 1,807 (90%) reported ever using non-injection drugs at their first eligible visit. These increased to 1,799 (90%) and 1,879 (94%) respectively at final visits. The median age of the included participants at the first included visit was 45 years (IQR 38, 51). Eligible participants were primarily male (70%) and white (70%). Most eligible participants were unemployed (70%) and had a monthly income less than $1500 CAD per month (78%). Few participants had end stage liver disease (9%) or a CD4 cell count less than 200 cells/μl (16%). Among those who experienced a fatal drug poisoning, individuals were more likely to be female, white, unhoused, a current smoker, have lower income, and have a detectable HIV viral load compared to those who did not experience a fatal drug poisoning event.

Table 2 – Baseline characteristics of participants at first eligible visit (N=1,998)

	Total (N=1998)	No drug poisoning within six months (N=1945)	Drug poisoning within six months (N=53)
Age (years)			
Median (IQR)	45 (38, 51)	45 (38, 51)	45 (41, 51)
Sex - Female	569 (29%)	552 (28%)	17 (32%)
Transgender	16 (1%)	15 (1%)	1 (2%)
Race / Ethnicity			
White	1391 (70%)	1351 (69%)	40 (75%)
Black	48 (2%)	46 (2%)	2 (4%)
Asian	26 (1%)	26 (1%)	0 (0%)
Hispanic/Latino	29 (1%)	27 (1%)	2 (4%)
Indigenous*	557 (28%)	545 (28%)	12 (23%)
High school education or higher	1491 (75%)	1453 (75%)	38 (72%)
Monthly income <$1500 CAD	1555 (78%)	1509 (78%)	46 (87%)
Unemployed	1400 (70%)	1362 (70%)	38 (72%)
Unhoused	228 (11%)	216 (11%)	12 (23%)
Current smoking	1568 (78%)	1521 (78%)	47 (89%)
Hazardous drinking	629 (31%)	613 (32%)	16 (30%)
CD4 levels <200 cells/µl	323 (16%)	311 (16%)	12 (23%)
HIV detectable viral load	467 (23%)	449 (23%)	18 (34%)
End stage liver disease	177 (9%)	169 (9%)	8 (15%)
STI ever	1135 (57%)	1106 (57%)	29 (55%)
Injection drug use ever	1764 (88%)	1713 (88%)	51 (96%)
Non-injection drug use ever	1807 (90%)	1756 (90%)	51 (96%)
Injection drug use current	983 (49%)	955 (49%)	28 (53%)
Non-injection drug use current	951 (48%)	921 (47%)	30 (57%)
Non-injection opioid use current	354 (18%)	338 (17%)	16 (30%)
Prescription opioids ever	661 (33%)	640 (33%)	21 (40%)
Addiction therapy current	421 (21%)	405 (21%)	16 (30%)

*Indigenous Peoples self-identified as First Nations (23%), Metis (5%), or Inuit (0.2%)
IQR = Interquartile range; STI = Sexually transmitted infection

Random forest model

Following the tuning procedure described above, the parameters used to fit the final random forest classifier were six for the number of variables randomly sampled at each node split, 500 for the number of classification trees in the random forest, and a minimum size of one for the terminal node. When applied to the entire sample, the model had excellent prediction with an AUC of 0.9965 (0.9941, 0.9988), an accuracy of 80.7%, a PPV of 1.5% and sensitivity of 100%. However, the model generated many false positives, as demonstrated by the low PPV (1.5%). As anticipated, OOB performance, which is considered to be more reliable, was poorer with an AUC of 0.61 (0.54, 0.68), an accuracy of 80.2%, a positive predictive value (PPV) of 0.4%, and a sensitivity of 28.3%.

Important predictors

We ranked variables by importance (see Figure 1) according to the mean decrease in classification accuracy. The top five most important variables were receiving addiction therapy in the past six months, sexually transmitted infection other than HIV or HCV (STI) ever, cigarette smoking in the past six months, being on prescription opioids ever, and non-injection opioid use in the past six months. However, the most important predictor only affected the mean accuracy by at most approximately seven points; that is, omitting the most important variable would misclassify, on average, approximately seven additional outcomes. The output of a GEE regression fitting drug poisoning as a function of the top predictors is listed in Table 2. Four of the five variables, addiction therapy (6 months), smoking (6 months), prescription opioids (ever), and non-injection opioid use (6 months) were associated with an increased odds of experiencing

the outcome; the confidence intervals for the associated odds ratios were wide and, with the exception of smoking, included the null value, thus precluding definitive conclusions.

Figure 1 – Variable importance plot of the top 10 variables

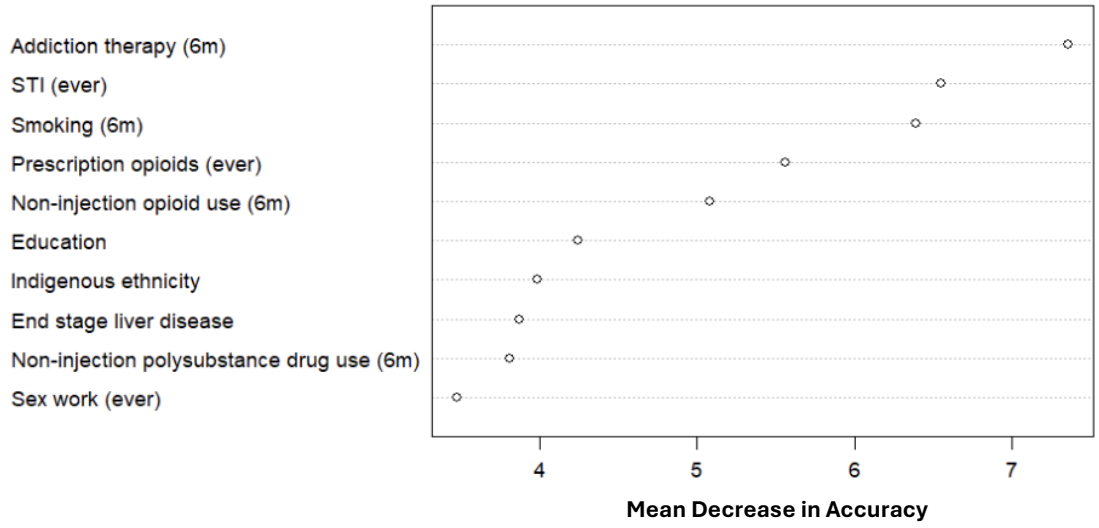

Table 3 – Association between the five most predictive variables, as measured by variable importance in the random forest, and drug poisoning deaths as captured via a GEE

Variable	Point estimate	Robust S.E.	Odds ratio (95% CI)
Addiction therapy	0.17	0.35	1.19 (0.60, 2.34)
STI (ever)	-0.08	0.28	0.93 (0.54, 1.60)
Cigarette smoking (6 months)	1.01	0.50	2.75 (1.03, 7.36)
Prescription opioids (ever)	0.36	0.32	1.43 (0.76, 2.68)
Non-injection opioid use (6 months)	0.55	0.33	1.73 (0.91, 3.29)

Sensitivity analysis

The model details of all sensitivity analyses can be found in Appendix 1. We ranked variables by importance for all sensitivity analyses and, to be consistent with the primary analysis, focused attention on the five top-ranked variables by importance. GEE regression models with a logistic

link and an exchangeable correlation structure were again fit using the five most predictive variables to provide a sense of the direction and magnitude of the association with drug poisoning; see Appendix 1. The results of all sensitivity analyses were similar to the main model. There was no improvement in precision, and no variable demonstrated strong predictive power.

There were 27 fatal drug poisoning events (8,093 non-events) for the analysis of the opioid era, 70 events (17,767 non-events) for the analysis including drug poisoning deaths up to one year, 90 events (17,747 non-events) for the analysis including unknown deaths, and 50 events (16,290 non-events) for the analysis of active users.

The predictors that consistently appeared in the top five most important predictors across all sensitivity analyses and the primary model were prescription opioids (ever) and addiction therapy (6 months). Both prescription opioids (ever) and addiction therapy (6 months) were associated with increased odds of experiencing an event; however, almost all 95% confidence intervals associated with these odds ratios were wide and crossed the null, reflecting the small numbers of drug poisonings in the CCC.

Discussion

We developed a random forest algorithm to predict drug poisoning deaths within six months of a cohort visit among people living with HIV-HCV co-infection using 40 candidate predictors drawn from socio-demographic, clinical, and behavioural data. Although the algorithm performed well (AUC = 0.9965) in-sample, the out-of-sample performance suggested results may not generalize well (OOB AUC = 0.61). The most important predictors of fatal drug

poisoning were recent (in last 6 months) addiction therapy, non-injection opioid use, and smoking and ever receiving prescription opioids or having an STI. Recent addiction therapy and ever receiving prescription opioids were also identified as predictors across all four sensitivity analyses. However, most of these predictors were weak. Our aim was to develop a tool for clinical use to flag at-risk patients and prevent drug poisoning deaths, however our results suggest that, at least within our cohort, drug poisoning deaths were difficult to predict. Poor predictive accuracy may have been due to the relatively small number of events. It is also possible that drug poisoning is driven more by external factors such as changing patterns of drug exposure and the increasingly toxic drug supply and their associated risk of drug poisoning over time than by individual patient-level factors.

Previous studies that have developed clinical prediction tools for drug poisonings, such as the ones identified in a review by Tseregounis & Henry, also used machine learning methods including random forest, deep neural network, and gradient boosting machine, and those with the best performance used clinical predictors (11). However, these studies had access to very large datasets (>25,000 individuals) with the majority including more than 100,000 individuals (11). As in our study, the prevalence of drug poisonings among these studies was quite low, ranging from 0.05% to 9.1% (11) however, given the larger datasets, the total number drug poisoning events was higher. It is possible we may have seen greater predictive accuracy with a larger dataset, as more data can improve a model's performance, especially when it is high quality data (23).

Although prediction was difficult, the top predictors we identified may provide clues to those most at risk. Certain predictors were consistent with previous studies. For example, prescription opioid use has been identified as an important predictor of drug poisoning events in several studies (24–27), including an increased risk associated with an increased dose (28).

Engaging in addiction therapy within the last 6 months might seem counterintuitive as an important predictor. Existing evidence suggests that retention in an addiction therapy program, such as opioid agonist therapy (OAT) reduces the risk of mortality (29). Such programs have seen considerable success in the context of harm reduction. However, studies have also found that in the period immediately following termination of a addiction treatment program, the risk of drug poisoning mortality is heightened (30–32). Additionally, individuals who access treatments such as OAT typically have more problematic substance use to begin with, namely opioid use disorder (OUD), and may be at heightened risk (33). A 2020 predictive modeling study identified the use of addiction treatment as an important predictor of experiencing an opioid drug poisoning (34). As it is not possible to determine when a CCC participant may have stopped an addiction therapy program following their last cohort visit, it is possible that it is not necessarily addiction therapy but perhaps stopping addiction therapy that is the important predictor.

Tobacco smoking and tobacco use have been identified as important predictors of drug poisoning (24,26,35). Smoking is highly prevalent among people who use drugs (36,37). A recent study investigating predictors of suicide or drug poisoning among individuals with any smoking exposure found that current smoking was associated with a risk of suicide/drug poisoning (38).

Although this study focused on both suicide and drug poisoning deaths, 56 of the 63 deaths were due to drug poisoning.

The existing literature on drug poisonings tends to focus on the injection route of administration due to its increased risks of harms, such as the risk of sexually transmitted and blood borne infection (STBBI) transmission (39). Our results suggest that the non-injection route should not be overlooked. In fact, the United States Centers for Disease Control and Prevention (CDC) reported that smoking was the leading route of drug use among drug poisoning deaths that occurred in 2022 (40). Non-injection prescription opioid use has also been identified as a risk factor for non-fatal drug poisonings among Canadian youth (41).

Finally, ever having experienced an STI appeared to be an important predictor in the random forest classifier, however the strength of the association was weak and the direction uncertain. It is possible that it is simply a marker of increased risk-taking behaviour.

Although the important predictors we identified in our analysis were also mentioned in several other studies, variables that were not available to us for this analysis such as a history of nonfatal drug poisonings and a diagnosis of opioid use disorder or substance use disorder have previously been identified as important predictors of drug poisonings across various studies (26,34,42–44).

Our study spanned a period of 20 years, a time during which drug use patterns have changed substantially in Canada (45) and within our cohort (46). Most previous studies have examined recent time periods of no longer than five years duration (11). Our model which restricted visits

to those in the "opioid era" (after 2016) performed best in the OOB sample (AUC 0.70). This may reflect a greater homogeneity of drug exposure and causes of drug poisoning, as drug use patterns of Canadians closer in time are more similar than those further apart (45). It may also reflect the changing landscape of toxic drug use patterns in Canada. Synthetic opioids such as fentanyl have completely transformed the illicit drug market, making it much deadlier (47). However, these did not become readily available in the street drug market until the early 2010s, with the first report of fentanyl in western Canada recorded in 2011 (47). This speaks to the potential randomness of drug poisonings in the cohort, as the composition of illicit street drugs is extremely variable. Other drugs such as stimulants, psychostimulants, and benzodiazepines have been on the rise in recent years (45,46,48). In addition, the drug poisoning epidemic has not impacted the entire country equally, with certain drugs being more prominent in western and northern Canada (46,47).

Strengths and limitations

Our study is the first to focus on predicting drug poisoning deaths within an HIV-HCV co-infected population. People living with HIV-HCV co-infection could benefit greatly from a predictive tool for drug poisonings as they are often more vulnerable and an important proportion engage in high-risk activities, such as drug use. Many of the strengths of this study lie in the data source. The participants in the CCC are generalizable to the HIV-HCV co-infected population in Canada that is linked to care (12). CCC participants are connected to care in a variety of settings across the country, including community-based clinics and university-based treatment programs, in both small and large urban centres. The richness and detail of the potential predictors available in the CCC is an additional strength. The method used, the random forest classifier, is also a

powerful and sophisticated machine learning algorithm that is considered to be accurate, relatively robust to outliers and noise, and is not overly prone to overfitting (15,17). Additionally, we conducted several sensitivity analyses which allowed us to explore model performance in different contexts.

This study has several limitations. Although drug poisoning was the most common known cause of death in our cohort affecting 5% of the participants included in these analyses, the number of events was relatively low which resulted in imbalanced data (53 events:17,784 non-events, or 0.3%). When the minority class is rare, a model can appear accurate when in reality it is prioritizing the majority class. We attempted to mitigate this imbalance by using a stratified random forest algorithm and under sampled the majority class. We opted for random undersampling, a popular choice to address data imbalance, as it is one of the simplest and most effective methods, and has shown to outperform more sophisticated methods (49). However, when undersampling the majority class, there is a loss of information that occurs as the model sees a smaller subset of the data. The data was not split into separate training and testing datasets and instead relied entirely on OOB estimates. Additionally, the algorithm was not validated on an external dataset. Certain variables in the CCC with limited data points had to be collapsed with other similar variables or removed entirely. The CCC does not collect data on non-fatal drug poisonings or OUD, which previous studies have found to be strongly associated with fatal drug poisonings (34,42,50). Including non-fatal poisonings would also have increased the number of events and thus increased our study power.

Conclusion

Despite a rich data set of socio-demographic, behavioural, and clinical predictors, we were unable to accurately predict six-month drug poisoning death in people living with HIV-HCV co-infection. While potential and plausible markers of vulnerability, such as recent drug treatment and certain recent drug exposures may signal increased risk, our model performance did not generalize well out of sample suggesting that drug poisonings may be driven more by external factors such an increasingly toxic and unpredictable drug supply. Future prediction studies would benefit from using a larger dataset in the recent time period to reduce heterogenicity in drug exposures. In the meantime, redoubling efforts to reduce harms from toxic drug exposures such as drug testing and drug purity kits, and public education is essential.

Appendix 1

Table 1 – Predictive performance of random forest for sensitivity analyses (OOB)

Sensitivity analysis	AUC (95% CI)	Accuracy	PPV	Sensitivity
One year	0.64 (0.58, 0.70)	80.2%	0.6%	28.6%
Opioid era	0.70 (0.59, 0.81)	83.8%	1.1%	51.9%
Active users	0.62 (0.56, 0.69)	79.4%	0.5%	30.0%
Unknown deaths	0.60 (0.55, 0.65)	79.8%	0.6%	22.2%

Table 2 – Predictive performance of random forest for sensitivity analyses (entire sample)

Sensitivity analysis	AUC (95% CI)	Accuracy	PPV	Sensitivity
One year	0.9958 (0.9921, 0.9995)	80.9%	2.0%	100%
Opioid era	0.9790 (0.9610, 0.9969)	84.2%	2.0%	96.3%
Active users	0.9976 (0.9961, 0.9991)	79.9%	1.5%	100%
Unknown deaths	0.9977 (0.9962, 0.9991)	80.7%	2.6%	100%

Table 3 – Association between the five most predictive variables of all analyses (primary and the four sensitivity analyses), as measured by variable importance in the random forest, and drug poisoning deaths as captured via a GEE

Variable	Main model OR (95% CI)	One year OR (95% CI)	Opioid era OR (95% CI)	Active users OR (95% CI)	Unknown deaths OR (95% CI)
Prescription opioids (ever)	1.43 (0.76, 2.68)	1.56 (0.92, 2.62)	3.07 (1.20, 7.87)	1.24 (0.66, 2.34)	1.52 (0.97, 2.40)
Addiction therapy (6m)	1.19 (0.60, 2.34)	1.21 (0.67, 2.17)	1.32 (0.53, 3.27)	1.23 (0.61, 2.45)	1.34 (0.82, 2.20)
STI (ever)	0.93 (0.54, 1.60)	0.90 (0.56, 1.46)			1.13 (0.73, 1.75)
Non-injection opioid use (6m)	1.73 (0.91, 3.29)			1.77 (0.82, 3.85)	
Cigarette smoking (6m)	2.75 (1.03, 7.36)				
Non-injection polysubstance drug use (6m)		1.03 (0.51, 2.06)		1.04 (0.46, 2.37)	
Injection stimulant use (6m)			1.22 (0.50, 2.97)		1.41 (0.90, 2.21)
Injection polysubstance drug use (6m)		1.87 (0.96, 3.61)			

Education (high school)			0.60 (0.26, 1.37)		
Sex (male)			0.66 (0.29, 1.53)		
Sex work (ever)				1.81 (1.01, 3.26)	
Sexual orientation (heterosexual)					1.67 (0.95, 2.93)

Acknowledgements: We would like to acknowledge the participants of the Canadian Co-infection Cohort, the Co-Investigators, Brian Conway, Mark Hull, Alexander Wong, John Gill, Lisa Barrett, Corinna Mae Quan, Joseph Cox, Pierre Côté, Shariq Haider, Danielle Rouleau, Valérie Martel-Laferrière, Marie-Louise Vachon, Tammy Bourque, Sharon Walmsley, Curtis Cooper, Stephen Sanche, Katherine Plewes, Jim Young, Neora Pick, and the study coordinators and staff.

Contributions:

Conceptualization: MB Klein, EEM Moodie, and M Bedard

Formal Analysis: M Bedard

Investigation: MB Klein, EEM Moodie, and M Bedard

Methodology: M Bedard and EEM Moodie

Writing – Original Draft: M Bedard

Writing – Review & Editing: M Bedard, MB Klein, EEM Moodie, J Cox, J Gill, S Walmsley, C Cooper and V Martel-Laferrière

Conflict of Interest

MB: None. EEMM: None. JC: JC has received research funding from ViiV Healthcare, Merck, and Gilead, consulting fees from ViiV Healthcare and Gilead, and payment for lectures from Gilead. JG: JG serves as an ad hoc advisor on Canadian National HIV advisory boards of Merck, Gilead and ViiV. SW: SW reports grants for investigator-initiated studies from ViiV Healthcare, Merck and Gilead and consulting fees and speaking at CME events for ViiV Healthcare, Merck, Gilead and Jannsen. VML: VML reports research grants from Merck and Gilead and consulting

fee from AbbVie. CC: CC reports speaker's fees, advisor fees and program support from Gilead Sciences, AbbVie and ViiV Healthcare. MBK: MBK has received grants for investigator-initiated studies from AbbVie, Gilead and ViiV Healthcare and consulting fees from ViiV Healthcare, AbbVie and Gilead.

Funding: This work was supported by the Canadian Institutes of Health Research (CIHR; FDN-143270 and PJT-185954); Fonds de recherche du Québec-Santé (FRQS), Réseau sida/maladies infectieuses; and the CIHR Canadian HIV Trials Network (CTN222). MLB is supported by a studentship award from the McGill University Health Centre-Research Institute. MBK is supported by a Tier I Canada Research Chair. VML is supported by Clinical Research Scholar – Junior 2 program (FRQ-S).

Data sharing policy: The datasets generated and/or analysed during the study are not publicly available.

References

1. Chen JY, Feeney ER, Chung RT. HCV and HIV co-infection: mechanisms and management. Nat Rev Gastroenterol Hepatol. 2014 Jun;11(6):362–71.

2. Gobran ST, Ancuta P, Shoukry NH. A tale of two viruses: Immunological insights into HCV/HIV coinfection. Front Immunol. 2021 Aug 12;12:726419.

3. Chu C, Gomes T, Antoniou T, Wong WWL, Janjua N, Guertin JR, et al. The impact of expanded access to direct acting antivirals for Hepatitis C virus on patient outcomes in Canada. PLOS ONE. 2023 Aug 8;18(8):e0284914.

4. World Health Organization. World Health Organization. [cited 2024 Sep 4]. People who inject drugs. Available from: https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/populations/people-who-inject-drugs

5. Kronfli N, Bhatnagar SR, Hull MW, Moodie EEM, Cox J, Walmsley S, et al. Trends in cause-specific mortality in HIV–hepatitis C coinfection following hepatitis C treatment scale-up. AIDS Lond Engl. 2019 May 1;33(6):1013–22.

6. Health Canada. Canada's overdose crisis and the toxic illegal drug supply [Internet]. 2024 [cited 2024 Jul 16]. Available from: https://www.canada.ca/en/health-canada/services/opioids/overdose-crisis-toxic-illegal-drug-supply.html

7. Public Health Agency of. Modelling opioid-related deaths during the overdose crisis [Internet]. 2020 [cited 2024 Aug 29]. Available from: https://www.canada.ca/en/health-canada/services/opioids/data-surveillance-research/modelling.html

8. Public Health Agency of Canada. Opioid- and stimulant-related harms in Canada: Key findings — Canada.ca [Internet]. 2024 [cited 2024 Jun 3]. Available from: https://health-infobase.canada.ca/substance-related-harms/opioids-stimulants/

9. Habehh H, Gohel S. Machine learning in healthcare. Curr Genomics. 2021 Dec 16;22(4):291–300.

10. Javaid M, Haleem A, Pratap Singh R, Suman R, Rab S. Significance of machine learning in healthcare: Features, pillars and applications. Int J Intell Netw. 2022 Jan 1;3:58–73.

11. Tseregounis IE, Henry SG. Assessing opioid overdose risk: a review of clinical prediction models utilizing patient-level data. Transl Res. 2021 Aug 1;234:74–87.

12. Klein MB, Saeed S, Yang H, Cohen J, Conway B, Cooper C, et al. Cohort Profile: The Canadian HIV–Hepatitis C Co-infection Cohort Study. Int J Epidemiol. 2010 Oct 1;39(5):1162–9.

13. CMHA. CMHA National. 2018 [cited 2023 May 23]. Overdose prevention. Available from: https://cmha.ca/brochure/overdose-prevention/

14. CHIP. CHIP – Centre of Excellence for Health, Immunity and Infections [Internet]. [cited 2024 Mar 5]. Available from: https://chip.dk/Home

15. Belyadi H, Haghighat A. Chapter 5 - Supervised learning. In: Belyadi H, Haghighat A, editors. Machine Learning Guide for Oil and Gas Using Python [Internet]. Gulf Professional Publishing; 2021 [cited 2024 Jun 12]. p. 169–295. Available from: https://www.sciencedirect.com/science/article/pii/B9780128219294000044

16. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. Am J Epidemiol. 2019 Dec 31;188(12):2222–39.

17. Cutler A, Cutler DR, Stevens JR. Random forests. In: Zhang C, Ma Y, editors. Ensemble Machine Learning: Methods and Applications [Internet]. New York, NY: Springer; 2012 [cited 2024 Jul 30]. p. 157–75. Available from: https://doi.org/10.1007/978-1-4419-9326-7_5

18. Chen C, Liaw A, Breiman L. Using random forest to learn imbalanced data [Internet]. University of California Berkeley; 2004 Jul. Report No.: 666. Available from: https://statistics.berkeley.edu/tech-reports/666

19. Ye Y, Wu Q, Zhexue Huang J, Ng MK, Li X. Stratified sampling for feature subspace selection in random forests for high dimensional data. Pattern Recognit. 2013 Mar 1;46(3):769–87.

20. Breiman L, Cutler A. Package 'randomForest' [Internet]. 2022. Available from: https://cran.r-project.org/web/packages/randomForest/randomForest.pdf

21. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008 Nov 10;28:1–26.

22. R Core Team. R: A language and environment for statistical computing. [Internet]. Vienna: R Foundation for Statistical Computing; 2022. Available from: https://www.r-project.org/

23. Junqué de Fortuny E, Martens D, Provost F. Predictive modeling with big data: Is bigger really better? Big Data. 2013 Dec;1(4):215–26.

24. Dong X, Deng J, Hou W, Rashidian S, Rosenthal RN, Saltz M, et al. Predicting opioid overdose risk of patients with opioid prescriptions using electronic health records based on temporal deep learning. J Biomed Inform. 2021 Apr 1;116:103725.

25. Ferris LM, Saloner B, Krawczyk N, Schneider KE, Jarman MP, Jackson K, et al. Predicting opioid overdose deaths using Prescription Drug Monitoring Program data. Am J Prev Med. 2019 Dec 1;57(6):e211–7.

26. Glanz JM, Narwaney KJ, Mueller SR, Gardner EM, Calcaterra SL, Xu S, et al. Prediction model for two-year risk of opioid overdose among patients prescribed chronic opioid therapy. J Gen Intern Med. 2018 Oct;33(10):1646–53.

27. Ripperger M, Lotspeich SC, Wilimitis D, Fry CE, Roberts A, Lenert M, et al. Ensemble learning to predict opioid-related overdose using statewide prescription drug monitoring program and hospital discharge data in the state of Tennessee. J Am Med Inform Assoc. 2022 Jan 1;29(1):22–32.

28. Adewumi AD, Hollingworth SA, Maravilla JC, Connor JP, Alati R. Prescribed dose of opioids and overdose: A systematic review and meta-analysis of unintentional prescription opioid overdose. CNS Drugs. 2018 Feb 1;32(2):101–16.

29. Pearce LA, Min JE, Piske M, Zhou H, Homayra F, Slaunwhite A, et al. Opioid agonist treatment and risk of mortality during opioid overdose public health emergency: population based retrospective cohort study. BMJ. 2020 Mar 31;m772.

30. Davoli M, Bargagli AM, Perucci CA, Schifano P, Belleudi V, Hickman M, et al. Risk of fatal overdose during and after specialist drug treatment: the VEdeTTE study, a national multi-site prospective cohort study. Addiction. 2007;102(12):1954–9.

31. Krawczyk N, Mojtabai R, Stuart EA, Fingerhood M, Agus D, Lyons BC, et al. Opioid agonist treatment and fatal overdose risk in a state-wide US population receiving opioid use disorder services. Addiction. 2020;115(9):1683–94.

32. Santo T Jr, Clark B, Hickman M, Grebely J, Campbell G, Sordo L, et al. Association of opioid agonist treatment with all-cause mortality and specific causes of death among people with opioid dependence: A systematic review and meta-analysis. JAMA Psychiatry. 2021 Sep 1;78(9):979–93.

33. Government of British Columbia. Province of British Columbia. Province of British Columbia; 2018 [cited 2024 Aug 19]. Opioid use disorder - Diagnosis and management in primary care. Available from: https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/bc-guidelines/opioid-use-disorder#scope

34. Saloner B, Chang HY, Krawczyk N, Ferris L, Eisenberg M, Richards T, et al. Predictive modeling of opioid overdose using linked statewide medical and criminal justice data. JAMA Psychiatry. 2020 Nov 1;77(11):1155–62.

35. Dong X, Rashidian S, Wang Y, Hajagos J, Zhao X, Rosenthal RN, et al. Machine learning based opioid overdose prediction using electronic health records. AMIA Annu Symp Proc. 2020 Mar 4;2019:389–98.

36. Guydish J, Passalacqua E, Pagano A, Martínez C, Le T, Chun J, et al. An international systematic review of smoking prevalence in addiction treatment. Addiction. 2016;111(2):220–30.

37. Kalman D, Morissette SB, George TP. Co-morbidity of smoking in patients with psychiatric and substance use disorders. Am J Addict. 2005;14(2):106–23.

38. Adviento BA, Regan EA, Make BJ, Han MK, Foreman MG, Iyer AS, et al. Clinical markers associated with risk of suicide or drug overdose among individuals with smoking exposure: A longitudinal follow-up study of the COPDGene cohort. CHEST. 2023 Feb 1;163(2):292–302.

39. Novak SP, Kral AH. Comparing injection and non-injection routes of administration for heroin, methamphetamine, and cocaine uses in the United States. J Addict Dis. 2011;30(3):248–57.

40. Tanz LJ, Gladden RM, Dinwiddie AT, Miller KD, Broz D, Spector E, et al. Routes of drug use among drug overdose deaths — United States, 2020–2022. MMWR Morb Mortal Wkly Rep. 2024;73:124–30.

41. Mitra G, Wood E, Nguyen P, Kerr T, DeBeck K. Drug use patterns predict risk of non-fatal overdose among street-involved youth in a Canadian setting. Drug Alcohol Depend. 2015 Aug 1;153:135–9.

42. Krawczyk N, Eisenberg M, Schneider KE, Richards TM, Lyons BC, Jackson K, et al. Predictors of overdose death among high-risk emergency department patients with

substance-related encounters: A data linkage cohort study. Ann Emerg Med. 2020 Jan 1;75(1):1–12.

43. Thylstrup B, Seid AK, Tjagvad C, Hesse M. Incidence and predictors of drug overdoses among a cohort of >10,000 patients treated for substance use disorder. Drug Alcohol Depend. 2020 Jan 1;206:107714.

44. Wang L, Hong PJ, Jiang W, Rehman Y, Hong BY, Couban RJ, et al. Predictors of fatal and nonfatal overdose after prescription of opioids for chronic pain: a systematic review and meta-analysis of observational studies. CMAJ Can Med Assoc J. 2023 Oct 23;195(41):E1399–411.

45. Health Canada. Canadian Alcohol and Drugs Survey (CADS): summary of results for 2019 [Internet]. 2021 [cited 2024 Aug 30]. Available from: https://www.canada.ca/en/health-canada/services/canadian-alcohol-drugs-survey/2019-summary.html

46. Lanièce Delaunay C, Maheu-Giroux M, Marathe G, Saeed S, Martel-Laferrière V, Cooper CL, et al. Gaps in hepatitis C virus prevention and care for HIV-hepatitis C virus co-infected people who inject drugs in Canada. Int J Drug Policy. 2022 May;103:103627.

47. Belzak L, Halverson J. Evidence synthesis - The opioid crisis in Canada: a national perspective. Health Promot Chronic Dis Prev Can Res Policy Pract. 2018 Jun;38(6):224–33.

48. National Institute on Drug Abuse. National Institute on Drug Abuse (NIDA). 2024 [cited 2024 Aug 30]. Drug overdose deaths: Facts and figures. Available from: https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates

49. Bach M, Werner A, Palt M. The proposal of undersampling method for learning from imbalanced datasets. Procedia Comput Sci. 2019 Jan 1;159:125–34.

50. Lo-Ciganic WH, Donohue JM, Yang Q, Huang JL, Chang CY, Weiss JC, et al. Developing and validating a machine-learning algorithm to predict opioid overdose in Medicaid beneficiaries in two US states: a prognostic modelling study. Lancet Digit Health. 2022 Jun;4(6):e455–65.

4. DISCUSSION

4.1. Summary of findings

The objective of the manuscript presented in this thesis was to predict six-month drug poisoning deaths within the Canadian HIV-HCV co-infection cohort. To accomplish this, we used a random forest classifier, a robust machine learning method used for prediction. The in-sample assessment of the model's predictive performance, which predicted events using the entire dataset, performed well (AUC 0.9965). However, the OOB assessment, which predicted events in a withheld sample of the data, performed poorly (AUC 0.61). These results indicate that our findings likely would not generalize well outside of the CCC, as OOB results are considered more reliable (less prone to optimism) than whole sample metrics. Additionally, we conducted four sensitivity analyses to explore model performance in different scenarios, and like the primary model, the models performed better on the entire sample than on a withheld portion of the sample across all sensitivity analyses. After ranking the variables by importance, the top five most important variables in the primary model were: addiction therapy in the past six months, STI ever, smoking in the past six months, prescription opioids ever, and non-injection opioid use in the past six months. Addiction therapy (6m) and prescription opioids (ever) appeared in the top five important predictors across all sensitivity analyses. However, none of these variables were strongly predictive: the variable importance plot demonstrated that they minimally affected the overall accuracy and the results from the GEE of the top five important predictors showed wide 95% confidence intervals, preventing any definitive conclusions from being made.

Previous studies, although conducted among populations other than the HIV-HCV co-infected population, such as Canadian youth or adults with at least one opioid prescription, have also

found similar important predictors to the top five highlighted in our analysis: prescription opioids, smoking, non-injection opioid use, and addiction therapy (16,89–92). Of our top five important predictors, there were two that stood out. Addiction therapy is a critical aspect of harm reduction that aims to lessen the harms related to drug use, including fatalities (93). It was therefore initially a surprising finding that this should predict drug poisoning. However, the current literature indicates that although addiction therapy programs are very helpful as long as individuals are engaged in the programs, the risk of drug poisoning appears to be heightened immediately after discontinuing treatment (60,94,95). This could be due to a variety of factors, such as no longer being accustomed to the current street drugs. Although it is not possible to know if a participant in the CCC would have discontinued addiction therapy after their final visit, this is a possibility. It may be beneficial to more closely monitor individuals who are actively in an addiction therapy program or have recently completed one and develop tailored education regarding risk of drug use under these circumstances. Another predictor of note among the top five was history of an STI. Although it appeared as an important predictor in the primary model and two sensitivity analysis models, the direction of the association was unclear. In the primary model and the sensitivity analysis including drug poisoning deaths up to one year after a participant's visit, STI appeared to be protective (although the 95% confidence intervals associated with the odds ratios crossed the null in both cases). In the sensitivity analysis including unknown deaths as drug poisoning deaths, the association indicated heightened risk, although once again the 95% confidence intervals associated with the odds ratio crossed the null. It is possible that reporting an STI is a marker of risk-taking behaviour; this warrants additional investigation.

A narrative review by Tseregounis & Henry identified 12 studies that developed clinical prediction tools for opioid poisonings (88). The studies included in this review identified similar important predictors to the ones we found in our analysis (16,17,92,96,97). The findings of this review indicate that prediction of drug poisoning events is possible, particularly if focusing on a well-defined type of process (e.g., poisoning in a particular era or from a particular drug class) or when using a very large sample. Many of the included studies focused on a narrow time period (<5 years) and had access to large datasets (>25,000 people). This is very different from our analysis which focused on a period of 20 years (thus spanning multiple drug eras) and included roughly 2,000 individuals. It is possible that our model may have seen better results if we had access to a larger dataset, particularly if such a larger dataset could be limited to a shorter period of time. In fact, the model of the sensitivity analysis limiting only to visits after 2016 saw the best OOB performance with an AUC of 0.70 (95% CI 0.59, 0.81) and a high in-sample model performance with an AUC of 0.9790 (95% CI 0.9610, 0.9969). This slightly improved performance may be due to a greater homogeneity in drug poisoning events occurring closer in time. Drug use patterns have changed significantly over the past 20 years, particularly after the introduction of synthetic opioids (98,99). It is riskier than ever to use drugs and drug poisonings are occurring at an alarmingly high rate (14,99). The drastic change in drug use patterns and in the supply itself throughout the duration of our study period have resulted in difficulty in predicting drug poisoning events in the cohort.

Finally, the imbalance in our data was an obstacle that we faced in this analysis. Although imbalanced data can pose a challenge, there have been instances of prediction with imbalanced data that have obtained good accuracy. For example, Khalilia et al. used ensemble learning

methods to predict disease risks in imbalanced data (100). Of the eight disease categories they predicted, the most imbalanced category, breast cancer, had a prevalence of 1.66%. The authors used a repeated random sub-sampling approach, and using this technique with random forest, they predicted breast cancer with an AUC of 0.9123 (100). However, this analysis exhibited two key advantages relative to our situation: the authors had access to much larger dataset (8 million records) as well as highly predictive features as the authors used diagnosis categories as predictors for diseases status. Authors Paing et al. compared four sampling methods for imbalanced data in random forest: random oversampling, random undersampling, Tomek Link, and synthetic minority oversampling technique (SMOTE) (101). Of these options, SMOTE yielded the best performance among a variety of imbalanced datasets, generating AUROC (area under receiver operating curve) values over 90 in all but one dataset. However, SMOTE is not suitable for categorical data (102), which feature in the CCC. Extensions to address categorical data either need a mix of continuous and categorical variables (102) or are not available in R. For our analysis, we opted for random undersampling, as it is simple yet effective (103). It is a popular choice to address imbalanced data and is readily available in R. There are many methods to address imbalanced data, each with their lists of pros and cons; it is unclear if using a different method to address the imbalance in our data would have generated better results, however an investigation into the difficult methodological question was beyond the scope of this thesis.

4.2. Strengths and limitations

There were several strengths in this thesis. One of the greatest strengths was the richness and detail of potential predictors in the CCC. The CCC collects information on hundreds of clinical, socio-demographic, and behavioural variables. This data provided a strong base for the random

forest classifier. Additionally, the participants in the CCC are generalizable to the HIV-HCV co-infected population in Canada connected to care. Participants are connected to care in small and large urban centres across the country (20). An additional strength of this thesis is the use of the random forest classifier, due to its reputation as a robust machine learning algorithm (81,104). The manuscript in this thesis is also the first study to focus on predicting drug poisoning deaths within the HIV-HCV co-infected population. Although this population is a small subset of the population, they could benefit greatly from a clinical prediction tool for drug poisonings as they are typically at higher risk and more vulnerable. Finally, we conducted several sensitivity analyses which allowed us to explore the model's performance in various contexts, which contributed relevant information to the study's objective.

There were several limitations to this thesis. As noted above, although drug poisoning deaths are the most common recorded cause of death in the CCC, the number of events was relatively low. To address the imbalance, we opted to use a stratified random forest and under-sample the majority class. However, when undersampling the majority class, there is a loss of information that occurs because not all observations are being used. A loss of information also occurred as certain variables in the CCC with variability in observed values had to be collapsed or removed entirely. We mitigated the loss of variables as much as possible by combining similar variables, for example combining drugs considered stimulants together such as cocaine and amphetamines. In addition, to not further reduce the number of events, we opted not to split the data into testing and training datasets and instead relied only on the whole sample and OOB results. Although OOB is a valuable tool in the random forest algorithm, as it withholds a portion of the data itself to then predict on the withheld sample, splitting the data into testing and training datasets would

ensure the model is truly blinded from the test dataset and provide a more accurate assessment of generalizability; unfortunately, the low number of events precluded this option. Working with a larger dataset with a greater number of events would help to address this problem, as even if a larger dataset had a low prevalence of events, the raw number would be greater than our current number of events, making splitting more feasible. We also did not validate the model on an external dataset. Finally, as is the reality with self-reported data, it is possible that participants may not have answered truthfully on questionnaires. However, the long and open relationship that CCC participants have with staff minimizes this concern. Despite these obstacles, we uncovered some interesting clues for who might be at risk and mitigated the challenges the best we could.

4.3. Future directions

This thesis was an important first step towards predicting drug poisoning deaths within the HIV-HCV co-infected population in Canada. However, our model did not perform well out of sample, indicating that it may not generalize well to other data including real-world clinic settings. Future studies within the cohort should examine different classification methods or, alternatively, this problem may require larger datasets which could be obtained by combining information across several cohorts to increase the number of events.

The CCC should also consider starting to collect additional information that could be helpful to predict drug poisoning events, such as experiencing a non-fatal drug poisoning event and a diagnosis of opioid use disorder. These variables proved to be important predictors in previous studies (92,105,106) and would be easy to incorporate into questionnaires. Additionally, if non-

fatal and fatal drug poisonings were to be combined, this would increase the number of events, thus helping to address the imbalanced data.

As prediction did not perform as well as anticipated, it is also possible that future public health efforts should focus elsewhere: rather than attempting to identify the most at-risk individuals, perhaps general messaging to people living with HIV-HCV co-infection who use drugs or focusing efforts on drug testing would be more fruitful, particularly as the current street drug market is alarmingly volatile.

5. CONCLUSION

In this thesis we used a random forest classifier to predict six-month drug poisoning deaths among individuals with HIV-HCV co-infection in Canada using sociodemographic, clinical, and behavioural data. Our model performed well in-sample, however, it performed poorly out of sample, indicating that our results may not generalize well. Although we identified certain important predictors, these results must be interpreted with caution. This was an important first attempt to develop a tool that could be deployed in clinical settings to reduce drug-related deaths within the HIV-HCV co-infected population. However, future efforts would benefit from using larger datasets, and within the cohort, amplifying drug poisoning prevention efforts such as drug testing and drug purity kits, or messaging to all drug users in the cohort.

6. REFERENCES

1. WHO. World Health Organization. 2024 [cited 2024 Aug 6]. HIV. Available from: https://www.who.int/data/gho/data/themes/hiv-aids

2. Public Health Agency of Canada. Estimates of HIV incidence, prevalence and Canada's progress on meeting the 90-90-90 HIV targets, 2020 [Internet]. Public Health Agency of Canada; 2022 Jul [cited 2024 Jun 24] p. 1–36. Available from: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/estimates-hiv-incidence-prevalence-canada-meeting-90-90-90-targets-2020.html

3. Public Health Agency of Canada. HIV and AIDS: Symptoms and treatment [Internet]. 2020 [cited 2024 Mar 5]. Available from: https://www.canada.ca/en/public-health/services/diseases/hiv-aids.html

4. Kemnic TR, Gulick PG. HIV antiretroviral therapy. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 [cited 2024 Aug 7]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK513308/

5. Public Health Agency of Canada. Pathogen safety data sheets: Infectious substances – human immunodeficiency virus (HIV) [Internet]. 2016 [cited 2024 Mar 7]. Available from: https://www.canada.ca/en/public-health/services/laboratory-biosafety-biosecurity/pathogen-safety-data-sheets-risk-assessment/human-immunodeficiency-virus.html

6. Public Health Agency of Canada. Government of Canada. 2024 [cited 2024 Aug 6]. Canada's progress towards ending the HIV epidemic. Available from: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/canada-progress-towards-ending-hiv-epidemic.html

7. Public Health Agency of Canada. Hepatitis C: Symptoms and treatment [Internet]. 2021 [cited 2024 Mar 8]. Available from: https://www.canada.ca/en/public-health/services/diseases/hepatitis-c.html

8. World Health Organization. Hepatitis C [Internet]. 2024 [cited 2024 Mar 8]. Available from: https://www.who.int/news-room/fact-sheets/detail/hepatitis-c

9. CATIE. CATIE - Canada's source for HIV and hepatitis C information. 2023 [cited 2024 Aug 6]. Hepatitis C treatment. Available from: https://www.catie.ca/hepatitis-c-an-in-depth-guide/hepatitis-c-treatment

10. Arends JE, Lieveld FI, Boeijen LL, de Kanter CTMM, van Erpecum KJ, Salmon D, et al. Natural history and treatment of HCV/HIV coinfection: Is it time to change paradigms? J Hepatol. 2015 Nov 1;63(5):1254–62.

11. Álvarez-Álvarez B, Prieto-Pérez L, de la Cuadra-Grande A, Casado MÁ, Cabello Úbeda A, Al-Hayani AW, et al. The era of DAAs: Assessing the patients' characteristics, clinical impact, and emergence of comorbidities in HIV/HCV-coinfected versus HIV-infected individuals. J Clin Med. 2024 Jan;13(13):3936.

12. CATIE. CATIE - Canada's source for HIV and hepatitis C information. 2024 [cited 2024 Aug 7]. Increased risk of death in Canadian study of people who fall out of care for HIV and hepatitis C. Available from: https://www.catie.ca/catie-news/increased-risk-of-death-in-canadian-study-of-people-who-fall-out-of-care-for-hiv-and

13. Hernandez MD, Sherman KE. HIV/HCV coinfection natural history and disease progression, a review of the most recent literature. Curr Opin HIV AIDS. 2011 Nov;6(6):478–82.

14. Public Health Agency of Canada. Opioid- and stimulant-related harms in Canada: Key findings [Internet]. 2024 [cited 2024 Apr 3]. Available from: https://health-infobase.canada.ca/substance-related-harms/opioids-stimulants/

15. World Health Organization. World Health Organization. [cited 2024 Sep 4]. People who inject drugs. Available from: https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/populations/people-who-inject-drugs

16. Ferris LM, Saloner B, Krawczyk N, Schneider KE, Jarman MP, Jackson K, et al. Predicting opioid overdose deaths using Prescription Drug Monitoring Program data. Am J Prev Med. 2019 Dec 1;57(6):e211–7.

17. Dong X, Rashidian S, Wang Y, Hajagos J, Zhao X, Rosenthal RN, et al. Machine learning based opioid overdose prediction using electronic health records. AMIA Annu Symp Proc. 2020 Mar 4;2019:389–98.

18. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science. 2015 Jul 17;349(6245):255–60.

19. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J. 2019 Jun;6(2):94–8.

20. Klein MB, Saeed S, Yang H, Cohen J, Conway B, Cooper C, et al. Cohort Profile: The Canadian HIV–Hepatitis C Co-infection Cohort Study. Int J Epidemiol. 2010 Oct 1;39(5):1162–9.

21. Public Health Agency of Canada. Sexually transmitted and blood borne infections (STBBI) prevention guide [Internet]. 2021 [cited 2024 Mar 5]. Available from: https://www.canada.ca/en/public-health/services/infectious-diseases/sexual-health-sexually-transmitted-infections/canadian-guidelines/stbbi-prevention-guide.html

22. Centers for Disease Control and Prevention. Centers for Disease Control and Prevention. 2021 [cited 2024 May 2]. Opportunistic infections. Available from: https://www.cdc.gov/hiv/basics/livingwithhiv/opportunisticinfections.html

23. Ibrahim A, Chattaraj A, Iqbal Q, Anjum A, Rehman MEU, Aijaz Z, et al. Pneumocystis jiroveci pneumonia: A review of management in human immunodeficiency virus (HIV) and non-HIV immunocompromised patients. Avicenna J Med. 2023 Mar 24;13(1):23–34.

24. Basavaraju A. Toxoplasmosis in HIV infection: An overview. Trop Parasitol. 2016;6(2):129–35.

25. World Health Organization. HIV and AIDS [Internet]. 2023 [cited 2024 Mar 7]. Available from: https://www.who.int/news-room/fact-sheets/detail/hiv-aids

26. The Global Fund. Key populations [Internet]. 2023 [cited 2024 Mar 8]. Available from: https://www.theglobalfund.org/en/key-populations/

27. CATIE. CATIE - Canada's source for HIV and hepatitis C information. 2023 [cited 2024 Mar 7]. HIV testing technologies. Available from: https://www.catie.ca/hiv-testing-technologies

28. Ontario HIV Treatment Network. Strategies to link people with undiagnosed HIV infection to HIV testing, care, and prevention services [Internet]. 2019 [cited 2024 Jul 24]. Available from: https://www.ohtn.on.ca/rapid-response-strategies-to-link-people-with-undiagnosed-hiv-infection-to-hiv-testing-care-and-prevention-services/

29. Minichiello A, Swab M, Chongo M, Marshall Z, Gahagan J, Maybank A, et al. HIV point-of-care testing in Canadian settings: A scoping review. Front Public Health. 2017 Apr 18;5:76.

30. Huynh K, Kahwaji CI. HIV testing. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 [cited 2024 Jul 24]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK482145/

31. Lo T, Fang CT, Lee YY, Shih CC, Chu FY, Wang JD. Early HIV diagnosis enhances quality-adjusted life expectancy of men who have sex with men living with HIV: A population-based cohort study in Taiwan. J Microbiol Immunol Infect. 2024 Feb 1;57(1):85–96.

32. Government of Canada. HIV: Undetectable = Untransmittable (U=U) [Internet]. 2024 [cited 2024 Jul 25]. Available from: https://www.canada.ca/en/services/health/campaigns/hiv-aids.html

33. CATIE. CATIE - Canada's source for HIV and hepatitis C information. 2024 [cited 2024 Mar 7]. HIV basics. Available from: https://www.catie.ca/essentials/hiv-basics

34. UNAIDS. UNAIDS. 2022 [cited 2024 Mar 7]. Global HIV & AIDS statistics — Fact sheet. Available from: https://www.unaids.org/en/resources/fact-sheet

35. CATIE. CATIE - Canada's source for HIV and hepatitis C information. 2024 [cited 2024 Mar 8]. The epidemiology of HIV in Canada. Available from: https://www.catie.ca/the-epidemiology-of-hiv-in-canada

36. UNAIDS. Understanding fast-track: Accelerating sction to rnd the AIDS epidemic by 2030 [Internet]. 2015 [cited 2024 May 10]. Available from:

https://www.unaids.org/sites/default/files/media_asset/201506_JC2743_Understanding_FastTrack_en.pdf

37. Maheu-Giroux M, Mishra S. Evidence with 95-95-95 that ambitious is feasible. Lancet HIV. 2024 Apr 1;11(4):e203–4.

38. UNAIDS. The path that ends AIDS: UNAIDS Global AIDS Update 2023. 2023 [cited 2024 May 15]; Available from: https://www.unaids.org/sites/default/files/media_asset/2023-unaids-global-aids-update_en.pdf

39. Public Health Agency of Canada. Pathogen safety data sheets: Infectious substances – hepatitis C virus [Internet]. 2010 [cited 2024 Mar 8]. Available from: https://www.canada.ca/en/public-health/services/laboratory-biosafety-biosecurity/pathogen-safety-data-sheets-risk-assessment/hepatitis-c-virus.html

40. Ashfaq UA, Javed T, Rehman S, Nawaz Z, Riazuddin S. An overview of HCV molecular biology, replication and immune responses. Virol J. 2011 Apr 11;8(1):161.

41. CATIE. CATIE - Canada's source for HIV and hepatitis C information. 2023 [cited 2024 Mar 8]. Hepatitis C: An in-depth guide. Available from: https://www.catie.ca/hepatitis-c-an-in-depth-guide

42. Falade-Nwulia O, Sulkowski MS, Merkow A, Latkin C, Mehta SH. Understanding and addressing hepatitis C reinfection in the oral direct acting antiviral era. J Viral Hepat. 2018 Mar;25(3):220–7.

43. Lynch EN, Russo FP. Outcomes and follow-up after hepatitis C eradication with direct-acting antivirals. J Clin Med. 2023 Mar 12;12(6):2195.

44. Mendizabal M, Piñero F, Ridruejo E, Herz Wolff F, Anders M, Reggiardo V, et al. Disease progression in patients with hepatitis C virus infection treated with direct-acting antiviral agents. Clin Gastroenterol Hepatol. 2020 Oct 1;18(11):2554-2563.e3.

45. Kalidindi Y, Jung J, Feldman R, Riley T III. Association of direct-acting antiviral treatment with mortality among Medicare beneficiaries with hepatitis C. JAMA Netw Open. 2020 Jul 21;3(7):e2011055.

46. CATIE. CATIE - Canada's source for HIV and hepatitis C information. 2024 [cited 2024 Mar 8]. The epidemiology of hepatitis C in Canada. Available from: https://www.catie.ca/the-epidemiology-of-hepatitis-c-in-canada-0

47. World Health Organization. Global health sector strategy on viral hepatitis 2016-2021. Towards ending viral hepatitis [Internet]. Geneva: World Health Organization; 2016. Available from: https://iris.who.int/handle/10665/246177

48. Platt L, Easterbrook P, Gower E, McDonald B, Sabin K, McGowan C, et al. Prevalence and burden of HCV co-infection in people living with HIV: a global systematic review and meta-analysis. Lancet Infect Dis. 2016 Jul 1;16(7):797–808.

49. Klein M, Rollet K, Saeed S, Cox J, Potter M, Cohen J, et al. HIV and hepatitis C virus coinfection in Canada: challenges and opportunities for reducing preventable morbidity and mortality. HIV Med. 2013;14(1):10–20.

50. Marino A, Zafarana G, Ceccarelli M, Cosentino F, Moscatt V, Bruno G, et al. Immunological and clinical impact of DAA-mediated HCV eradication in a cohort of HIV/HCV coinfected patients: Monocentric Italian experience. Diagnostics. 2021 Dec 11;11(12):2336.

51. Bruno G, Saracino A, Scudeller L, Fabrizio C, Dell'Acqua R, Milano E, et al. HCV mono-infected and HIV/HCV co-infected individuals treated with direct-acting antivirals: to what extent do they differ? Int J Infect Dis. 2017 Sep 1;62:64–71.

52. Kronfli N, Bhatnagar SR, Hull MW, Moodie EEM, Cox J, Walmsley S, et al. Trends in cause-specific mortality in HIV–hepatitis C coinfection following hepatitis C treatment scale-up. AIDS Lond Engl. 2019 May 1;33(6):1013–22.

53. St-Jean M, Dong X, Tafessu H, Moore D, Honer WG, Vila-Rodriguez F, et al. Overdose mortality is reducing the gains in life expectancy of antiretroviral-treated people living with HIV in British Columbia, Canada. Int J Drug Policy. 2021 Oct 1;96:103195.

54. CMHA. CMHA National. 2018 [cited 2023 May 23]. Overdose prevention. Available from: https://cmha.ca/brochure/overdose-prevention/

55. Government of British Columbia. BC Gov News. 2016 [cited 2024 Jun 20]. Provincial health officer declares public health emergency. Available from: https://news.gov.bc.ca/releases/2016HLTH0026-000568

56. Government of British Columbia. BC Gov News. 2024 [cited 2024 Jun 20]. Premier's, minister's, provincial health officer's statements on the eighth anniversary of toxic-drug public-health emergency. Available from: https://news.gov.bc.ca/releases/2024MMHA0020-000552

57. Public Health Agency of Canada. Joint statement from the co-chairs of the special advisory committee on the epidemic of opioid overdoses – Latest national data on substance-related harms [Internet]. 2024 [cited 2024 Jun 24]. Available from: https://www.canada.ca/en/public-health/news/2024/03/joint-statement-from-the-co-chairs-of-the-special-advisory-committee-on-the-epidemic-of-opioid-overdoses--latest-national-data-on-substance-related.html

58. Commonwealth of Massachusetts. Opioid overdose risk factors [Internet]. [cited 2024 Apr 3]. Available from: https://www.mass.gov/info-details/opioid-overdose-risk-factors

59. Green TC, McGowan SK, Yokell MA, Pouget ER, Rich JD. HIV infection and risk of overdose: a systematic review and meta-analysis. AIDS Lond Engl. 2012 Feb 20;26(4):403–17.

60. Krawczyk N, Mojtabai R, Stuart EA, Fingerhood M, Agus D, Lyons BC, et al. Opioid agonist treatment and fatal overdose risk in a state-wide US population receiving opioid use disorder services. Addiction. 2020;115(9):1683–94.

61. Canadian Centre on Substance Use and Addiction. Changes related to COVID-19 in the illegal drug supply and access to services, and resulting health harms [Internet]. 2020 May. Available from: https://www.ccsa.ca/sites/default/files/2020-05/CCSA-COVID-19-CCENDU-Illegal-Drug-Supply-Alert-2020-en.pdf

62. Peppin JF, Raffa RB, Schatman ME. The polysubstance overdose-death crisis. J Pain Res. 2020 Dec 15;13:3405–8.

63. Centers for Disease Control and Prevention. Stop Overdose. 2022 [cited 2024 Apr 30]. Polysubstance use facts. Available from: https://www.cdc.gov/stopoverdose/polysubstance-use/index.html

64. Caudarella A, Dong H, Milloy MJ, Kerr T, Wood E, Hayashi K. Non-fatal overdose as a risk factor for subsequent fatal overdose among people who inject drugs. Drug Alcohol Depend. 2016 May 1;162:51–5.

65. Lipato T, Terplan M. Risk factors for opioid overdose. Curr Treat Options Psychiatry. 2018 Sep 1;5(3):323–33.

66. Schiavon S, Hodgin K, Sellers A, Word M, Galbraith JW, Dantzler J, et al. Medical, psychosocial, and treatment predictors of opioid overdose among high risk opioid users. Addict Behav. 2018 Nov 1;86:51–5.

67. Tomko C, Schneider KE, Rouhani S, Urquhart GJ, Nyeong Park J, Morris M, et al. Identifying pathways to recent non-fatal overdose among people who use opioids non-medically: How do psychological pain and unmet mental health need contribute to overdose risk? Addict Behav. 2022 Apr 1;127:107215.

68. Smolina K, Crabtree A, Chong M, Zhao B, Park M, Mill C, et al. Patterns and history of prescription drug use among opioid-related drug overdose cases in British Columbia, Canada, 2015–2016. Drug Alcohol Depend. 2019 Jan 1;194:151–8.

69. Cunningham CO. Opioids and HIV infection: From pain management to addiction treatment. Top Antivir Med. 2018 Apr;25(4):143–6.

70. Touesnard N, Brothers TD, Bonn M, Edelman EJ. Overdose deaths and HIV infections among people who use drugs: shared determinants and integrated responses. Expert Rev Anti Infect Ther. 2022 Aug 3;20(8):1061–5.

71. Samji H, Yu A, Wong S, Wilton J, Binka M, Alvarez M, et al. Drug-related deaths in a population-level cohort of people living with and without hepatitis C virus in British Columbia, Canada. Int J Drug Policy. 2020 Dec;86:102989.

72. Kim D, Alshuwaykh OS, Cholankeril G, Wong RJ, Ahmed A. Trends in mortality in hepatitis C infection and alcoholic liver disease based on drug overdose in the United States. J Viral Hepat. 2021;28(2):436–9.

73. Korona-Bailey J, Riley Saint S, Sizemore L, Wingate H, Shoup P, Hawes A, et al. Prevalence of hepatitis C virus among fatal drug overdoses in Tennessee: an analysis using 2019–2020 Tennessee state unintentional drug overdose reporting system data. Ann Epidemiol. 2023 Apr 1;80:1–8.

74. Perlman DC, Jordan AE. The syndemic of opioid misuse, overdose, HCV, and HIV: Structural-level causes and interventions. Curr HIV/AIDS Rep. 2018 Apr 1;15(2):96–112.

75. Canadian Mental Health Association. Canadian Mental Health Association. [cited 2024 Jun 18]. Harm reduction. Available from: https://ontario.cmha.ca/harm-reduction/

76. Government of Canada. Federal actions on the overdose crisis [Internet]. 2020 [cited 2024 May 15]. Available from: https://www.canada.ca/en/health-canada/services/opioids/federal-actions/overview.html

77. Palmateer N, Hamill V, Bergenstrom A, Bloomfield H, Gordon L, Stone J, et al. Interventions to prevent HIV and hepatitis C among people who inject drugs: Latest evidence of effectiveness from a systematic review (2011 to 2020). Int J Drug Policy. 2022 Nov 1;109:103872.

78. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. Am J Epidemiol. 2019 Dec 31;188(12):2222–39.

79. Sarker IH. Machine learning: Algorithms, real-world applications and research directions. SN Comput Sci. 2021 Mar 22;2(3):160.

80. Badillo S, Banfai B, Birzele F, Davydov II, Hutchinson L, Kam-Thong T, et al. An introduction to machine learning. Clin Pharmacol Ther. 2020;107(4):871–85.

81. Belyadi H, Haghighat A. Chapter 5 - Supervised learning. In: Belyadi H, Haghighat A, editors. Machine Learning Guide for Oil and Gas Using Python [Internet]. Gulf Professional Publishing; 2021 [cited 2024 Jun 12]. p. 169–295. Available from: https://www.sciencedirect.com/science/article/pii/B9780128219294000044

82. Habehh H, Gohel S. Machine learning in healthcare. Curr Genomics. 2021 Dec 16;22(4):291–300.

83. Shen J, Zhang CJP, Jiang B, Chen J, Song J, Liu Z, et al. Artificial intelligence versus clinicians in disease diagnosis: systematic review. JMIR Med Inform. 2019 Aug 16;7(3):e10010.

84. Bertsimas D, Wiberg H. Machine learning in oncology: Methods, applications, and challenges. JCO Clin Cancer Inform. 2020 Dec;(4):885–94.

85. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. PLOS ONE. 2018 Jul 20;13(7):e0201016.

86. Daneshjou R, Vodrahalli K, Novoa RA, Jenkins M, Liang W, Rotemberg V, et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. Sci Adv. 2022 Aug 12;8(32):eabq6147.

87. Rubinger L, Gazendam A, Ekhtiari S, Bhandari M. Machine learning and artificial intelligence in research and healthcare. Injury. 2023 May 1;54:S69–73.

88. Tseregounis IE, Henry SG. Assessing opioid overdose risk: a review of clinical prediction models utilizing patient-level data. Transl Res. 2021 Aug 1;234:74–87.

89. Dong X, Deng J, Hou W, Rashidian S, Rosenthal RN, Saltz M, et al. Predicting opioid overdose risk of patients with opioid prescriptions using electronic health records based on temporal deep learning. J Biomed Inform. 2021 Apr 1;116:103725.

90. Ripperger M, Lotspeich SC, Wilimitis D, Fry CE, Roberts A, Lenert M, et al. Ensemble learning to predict opioid-related overdose using statewide prescription drug monitoring program and hospital discharge data in the state of Tennessee. J Am Med Inform Assoc. 2022 Jan 1;29(1):22–32.

91. Mitra G, Wood E, Nguyen P, Kerr T, DeBeck K. Drug use patterns predict risk of non-fatal overdose among street-involved youth in a Canadian setting. Drug Alcohol Depend. 2015 Aug 1;153:135–9.

92. Saloner B, Chang HY, Krawczyk N, Ferris L, Eisenberg M, Richards T, et al. Predictive modeling of opioid overdose using linked statewide medical and criminal justice data. JAMA Psychiatry. 2020 Nov 1;77(11):1155–62.

93. Pearce LA, Min JE, Piske M, Zhou H, Homayra F, Slaunwhite A, et al. Opioid agonist treatment and risk of mortality during opioid overdose public health emergency: population based retrospective cohort study. BMJ. 2020 Mar 31;m772.

94. Davoli M, Bargagli AM, Perucci CA, Schifano P, Belleudi V, Hickman M, et al. Risk of fatal overdose during and after specialist drug treatment: the VEdeTTE study, a national multi-site prospective cohort study. Addiction. 2007;102(12):1954–9.

95. Santo T Jr, Clark B, Hickman M, Grebely J, Campbell G, Sordo L, et al. Association of opioid agonist treatment with all-cause mortality and specific causes of death among people with opioid dependence: A systematic review and meta-analysis. JAMA Psychiatry. 2021 Sep 1;78(9):979–93.

96. Chang HY, Ferris L, Eisenberg M, Krawczyk N, Schneider KE, Lemke K, et al. The impact of various risk assessment time frames on the performance of opioid overdose forecasting models. Med Care. 2020 Nov;58(11):1013.

97. Glanz JM, Narwaney KJ, Mueller SR, Gardner EM, Calcaterra SL, Xu S, et al. Prediction model for two-year risk of opioid overdose among patients prescribed chronic opioid therapy. J Gen Intern Med. 2018 Oct;33(10):1646–53.

98. National Institute on Drug Abuse. National Institute on Drug Abuse (NIDA). 2024 [cited 2024 Aug 30]. Drug overdose deaths: Facts and figures. Available from: https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates

99. Health Canada. Canadian Alcohol and Drugs Survey (CADS): summary of results for 2019 [Internet]. 2021 [cited 2024 Aug 30]. Available from: https://www.canada.ca/en/health-canada/services/canadian-alcohol-drugs-survey/2019-summary.html

100. Khalilia M, Chakraborty S, Popescu M. Predicting disease risks from highly imbalanced data using random forest. BMC Med Inform Decis Mak. 2011 Jul 29;11(1):51.

101. Paing MP, Pintavirooj C, Tungjitkusolmun S, Choomchuay S, HAMAMOTO K. Comparison of sampling methods for imbalanced data classification in random forest. In: 2018 11th Biomedical Engineering International Conference (BMEiCON) [Internet]. 2018 [cited 2024 Aug 16]. p. 1–5. Available from: https://ieeexplore.ieee.org/abstract/document/8609946?casa_token=XzJ2L9MWPjsAAAAA:htwQQKsYOG1DtDhAnh2yzOuDhhikJ07J4gNXWeyIuYoDYsc1JRqwVDCFkMDy6T7gJ2728EwI

102. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res. 2002 Jun 1;16:321–57.

103. Bach M, Werner A, Palt M. The proposal of undersampling method for learning from imbalanced datasets. Procedia Comput Sci. 2019 Jan 1;159:125–34.

104. Cutler A, Cutler DR, Stevens JR. Random forests. In: Zhang C, Ma Y, editors. Ensemble Machine Learning: Methods and Applications [Internet]. New York, NY: Springer; 2012 [cited 2024 Jul 30]. p. 157–75. Available from: https://doi.org/10.1007/978-1-4419-9326-7_5

105. Krawczyk N, Eisenberg M, Schneider KE, Richards TM, Lyons BC, Jackson K, et al. Predictors of overdose death among high-risk emergency department patients with substance-related encounters: A data linkage cohort study. Ann Emerg Med. 2020 Jan 1;75(1):1–12.

106. Lo-Ciganic WH, Donohue JM, Yang Q, Huang JL, Chang CY, Weiss JC, et al. Developing and validating a machine-learning algorithm to predict opioid overdose in Medicaid beneficiaries in two US states: a prognostic modelling study. Lancet Digit Health. 2022 Jun;4(6):e455–65.