



PREDICTING THE PRESENCE OF CLINICALLY SIGNIFICANT DEPRESSIVE SYMPTOMS IN THE HIV-HEPATITIS C CO-INFECTED POPULATION IN CANADA USING SUPERVISED MACHINE LEARNING

Gayatri Marathe¹, Erica Moodie¹, Marie-Josée Brouillette^{1,2}, Joseph Cox^{1,2}, Curtis Cooper³, Brian Conway⁴, Mark Hull⁵, Marie-Louise Vachon⁶, Sharon Walmsley⁷, Alexander Wong⁸, Marina Klein^{1,2}; Canadian Co-Infection Cohort

1. McGill University, Montreal, Quebec; 2. McGill University Health Center-Research Institute, Montreal, Quebec; 3. Department of Medicine, University of Ottawa, Ottawa, Ontario; 4. Vancouver Infectious Diseases Centre, Vancouver, British Columbia; 5. Centre for Excellence in HIV/AIDS, St. Paul's Hospital, Vancouver, British Columbia; 6. Centre Hospitalier de l'Université Laval, Quebec, Quebec; 7. Toronto General Hospital Research Institute, Toronto, Ontario; 8. Regina General Hospital, Saskatchewan Health Authority, Regina, Saskatchewan

No conflicts of interest to disclose in relation to this work.

Contact email: gayatri.marathe@mail.mcgill.ca

BACKGROUND AND OBJECTIVE

- HIV-Hepatitis C Virus (HCV) co-infection - Common modes of transmission.
- Depression - Most common associated neuropsychiatric disorder
 - Prevalence estimates: HIV: ~ 42 %¹ | Chronic HCV: ~ 24 %²
- The prevalence is suggested to be higher in HIV-HCV co-infected population.³
- Depression screening may not be a priority due to competing health concerns and other priorities in this population and depression underdiagnosis is an issue.
- Demographic, behavioural and clinical characteristics are associated with presence of depressive symptoms and this data may be routinely collected in clinical and research settings.

Objective

To develop algorithms to predict the presence of clinically significant depressive symptoms in the HIV-HCV co-infected population in Canada using patient data and supervised machine learning.

DATA SOURCES

Canadian Co-infection Cohort (CCC)

- Multicenter prospective cohort study with visits every 6 months - ongoing since 2003
- HIV-infected participants with evidence of HCV infection - 2009 participants as of February 2020.

Food Security and HIV-HCV co-infection study (FS sub-study)

- Sub-study initiated in 2012 until 2015, with visits integrated into the CCC - Total 725 participants.
- Depression screening was performed using Center for Epidemiologic Studies Depression Scale-10 (CES-D-10):
 - 10 item Likert scale, score range: 0-30
 - CES-D-10 classes (I/0) at score 10



CES-D-10 score ≥ 10 :

At risk for major depression with presence of clinically significant depressive symptoms

Analytical sample

- FS sub-study visits merged with corresponding CCC visit data (within a 3-month window).
- Excluded FS sub-study visits, i. with no CES-D-10 measures, and ii. with no corresponding CCC visit.

METHODS

- Predictors from 5 broad categories used (total number = 137):
 - Related to psychiatric illness, health related quality of life, demographic, behavioural and clinical characteristics.
- Supervised machine learning using Random Forest (RF) - Decision tree based method.
- Two RF algorithms were developed using the training data (70% subset):
 - Full algorithm** - all candidate predictors ($x=137$)
 - Reduced algorithm** - routinely available predictors based on expert opinion ($x=46$)
- Performance was evaluated using the testing data (30% subset):
 - Area under the ROC curve (AUC)
 - Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were estimated at the optimal probability threshold (maximum Youden index (J) = Sens + Spec - 1).
- Predictors important for CES-D-10 classification were identified using the RF importance metrics.

RESULTS AND CONCLUSION

- We included 1934 visits from 717 FS sub-study participants.
- Participants were predominantly male (73%), white (76%) and unemployed (73%).
- 53% participants presented with clinically significant depressive symptoms in the first visit.

Evaluation measure (95% CI)	Full Algorithm (x=137)	Reduced Algorithm (x=46)
AUC	0.86 (0.83-0.89)	0.77 (0.73-0.80)
Sensitivity	0.77 (0.72-0.82)	0.73 (0.67-0.78)
Specificity	0.77 (0.72-0.82)	0.69 (0.63-0.75)
PPV	0.77 (0.72-0.82)	0.71 (0.65-0.76)
NPV	0.77 (0.72-0.82)	0.72 (0.66-0.77)

- Employment, HIV clinical stage, revenue source, BMI, age, and education were among the 15 most important predictors in both algorithms.

Conclusion

- The algorithms show promise and overall the full algorithm performed better than the reduced.
- There is possible misclassification and external validation for generalizability beyond the CCC is needed.
- However, with lack of routine screening in standard-of-care, integration of such automated algorithms may be of great value to avoid missing individuals at risk for major depression.