

# Predictive variables for peripheral neuropathy in treated HIV type 1 infection revealed by machine learning

Wei Tu<sup>a,f,g</sup>, Erika Johnson<sup>b</sup>, Esther Fujiwara<sup>d</sup>, M. John Gill<sup>c,e</sup>,  
Linglong Kong<sup>a</sup> and Christopher Power<sup>b,d,e</sup>

**Objective:** Peripheral neuropathies (PNPs) in HIV-infected patients are highly debilitating because of neuropathic pain and physical disabilities. We defined prevalence and associated predictive variables for PNP subtypes in a cohort of persons living with HIV.

**Design:** Adult persons living with HIV in clinical care were recruited to a longitudinal study examining neurological complications.

**Methods:** Each patient was assessed for symptoms and signs of PNP with demographic, laboratory, and clinical variables. Univariate, multiple logistic regression and machine learning analyses were performed by comparing patients with and without PNP.

**Results:** Three patient groups were identified: PNP ( $n=111$ ) that included HIV-associated distal sensory polyneuropathy ( $n=90$ ) or mononeuropathy ( $n=21$ ), and non-neuropathy ( $n=408$ ). Univariate analyses showed multiple variables differed significantly between the non-neuropathy and PNP groups including age, estimated HIV type 1 (HIV-1) duration, education, employment, neuropathic pain, peak viral load, polypharmacy, diabetes, cardiovascular disorders, AIDS, and prior neurotoxic nucleoside antiretroviral drug exposure. Classification algorithms distinguished those with PNP, all with area under the receiver operating characteristic curve values of more than 0.80. Random forest models showed greater accuracy and area under the receiver operating characteristic curve values compared with the multiple logistic regression analysis. Relative importance plots showed that the foremost predictive variables of PNP were HIV-1 duration, peak plasma viral load, age, and low CD4<sup>+</sup> T-cell levels.

**Conclusion:** PNP in HIV-1 infection remains common affecting 21.4% of patients in care. Machine-learning models uncovered variables related to PNP that were undetected by conventional analyses, emphasizing the importance of statistical algorithmic approaches to understanding complex neurological syndromes.

Copyright © 2021 Wolters Kluwer Health, Inc. All rights reserved.

*AIDS* 2021, **35**:1785–1793

**Keywords:** antiretroviral neurotoxicity, comorbidity, distal sensory polyneuropathy, machine learning, mononeuropathy

---

<sup>a</sup>Department of Mathematical & Statistical Sciences, <sup>b</sup>Department of Medicine, University of Alberta, Edmonton, <sup>c</sup>Departments of Medicine, University of Calgary, Calgary, <sup>d</sup>Department of Psychiatry, University of Alberta, Edmonton, <sup>e</sup>Department of Microbiology, Immunology & Infectious Diseases, University of Calgary, Calgary, AB, <sup>f</sup>Department of Public Health Sciences, and <sup>g</sup>Canadian Cancer Trials Group, Queen's University, Kingston, ON, Canada.

Correspondence to Christopher Power, Department of Medicine (Neurology), University of Alberta, 6-11 Heritage Medical Research Centre, Edmonton, AB, Canada.

Tel: +1 780 407 1938; fax: +1 780 407 1984; e-mail: chris.power@ualberta.ca

Received: 22 October 2020; revised: 14 April 2021; accepted: 10 May 2021.

DOI:10.1097/QAD.0000000000002955

## Introduction

Peripheral nerve disorders have long been recognized as a feature of HIV type 1 (HIV-1) infection and are most apparent following progression to AIDS [1]. Indeed, the spectrum of peripheral neuropathies (PNPs) that occur during HIV/AIDS is broad, ranging from acute inflammatory demyelinating polyneuropathy (Guillain-Barre syndrome), distal sensory (axonal) polyneuropathy (DSP) to mononeuropathies (MNP) [2]. These disorders are often associated with neuropathic pain and physical disabilities including sensory loss, paresthesia, weakness, ataxia, and gait dysfunction as well as concurrent co-morbidities including diabetes, syphilis, and depression. Several of the early antiretroviral therapies (ARTs), including the so-called d-drugs, such as didanosine (ddI), zalcitabine (ddC), and stavudine (d4T) have been associated with a sensory axonal polyneuropathy, sometimes accompanied by neuropathic pain [3]. While these neurotoxic medications are no longer components of contemporary treatment, past exposure remains important. The underlying pathogenesis of PNPs associated with HIV/AIDS remains obscure although both clinical and experimental studies point to roles for mitochondrial injury [4], neurotrophin depletion [5] as well as direct cytopathic effects of viral proteins [6]. Nonetheless, both polyneuropathies and MNPs are common to other (animal) lentivirus infections [7,8], highlighting the importance of viral infection and the associated pathogenic effects in the occurrence of PNPs.

Modern ART has dramatically improved both the mortality and the morbidity from HIV-1 infection leading to a close to normal life expectancy for most persons living with HIV (PWH) [9]. Routine care for PWH today often involves management of co-morbidities such as cardiovascular disease, cancer, bone disease, and/or neurological disease. Advancing age and frailty underpins the emergence of many of these disorders [10]. Moreover, pain and different physical disabilities frequently also complicate these disorders. PNP is more common in all older persons and can be associated with neuropathic pain [11]. We hypothesized that given the complexity of factors contributing to PNPs, implementing diverse statistical approaches to a well defined dataset would yield new insights into the contributing variables. To gain a deeper understanding of the relative importance of clinical and demographic variables involved in PNP occurrence and its subtypes, we compared univariate, multiple logistic regression with machine learning analyses of this cohort. The prevalence and associated variables were assessed among adult PWHs with diverse ethnic backgrounds with controlled HIV-1 infection with or without PNP in a contemporary clinical setting in which patients received long-term clinical follow-up under universal healthcare.

## Methods and materials

### Patient cohort

All HIV-1 seropositive adult patients at the Southern Alberta (HIV) Clinic (SAC) in Calgary, Alberta from 2013 to 2019 were invited to enroll in a study assessing neurologic complications of HIV/AIDS during routine clinical care provided by a SAC physician, which included a general inquiry and physical examination. Exclusion criteria included non-fluency in English, less than 18 years of age, less than a grade 9 education, the presence of severe psychiatric (e.g. schizophrenia) or neurological disorders (e.g. brain tumors, strokes, epilepsy), history of brain damage/traumatic brain injury with loss of consciousness (>5 min), and uncorrected vision or hearing impairments [12–15]. At the time of recruitment and ongoing, all patients were routinely asked about symptoms of pain, sensory abnormalities or weakness, which if reported, prompted an examination by a neurologist to verify sensory or motor deficits. Patients were determined to have PNP (DSP or MNP) if two or more of the following criteria were present: first, diffuse or focal sensory symptoms including numbness, paresthesia, or neuropathic pain (e.g. continuous or intermittent, evoked or spontaneous dysesthesia, hyperalgesia, allodynia) with associated descriptors ('burning', 'stabbing'); second, abnormal sensory signs on physical exam such as bilateral reduced vibratory perception, glove-stocking sensory loss or focal sensory deficits; third, focal weakness; and/or fourth, decreased or absent ankle reflexes. The diagnosis and type of PNP was determined based on review of the reported symptoms, laboratory results, and physical examination [16–18]. The presence of neuropathic pain was predicated on subjects' reports of the above symptoms (e.g. dysesthesia, hyperalgesia, allodynia) together confirmation of frequency, duration and severity together with a plausible anatomic distribution [19,20]. MNPs were diagnosed in patients with a discernible anatomic localization and pattern (e.g. carpal tunnel, facial neuropathy, trigeminal neuropathy, or focal radiculopathy) [21]. Electromyography and nerve conduction studies were performed in select patients to verify MNPs and to exclude other neuromuscular disorders. Multiple variables were assessed including health-related quality of life (HQoL), number of hours of sleep per night, presence and severity of depressive symptoms (Patient Health Questionnaire assay, PHQ-9) and neurocognitive symptoms. The University of Calgary Ethics Committee (REB #-130615) approved the study and written consent was obtained from all patients.

### Study setting and design

The SAC serves all PWH (currently ~1800 persons) in HIV care in Southern Alberta (estimated 2020 total population, ~2.4 million) and is a multi-disciplinary clinic which opened in 1989. SAC offers regular clinical follow-up visits, laboratory investigations, and ART all at no cost to the patient. SAC has a multi-disciplinary team

including physicians, nurses, social workers, dieticians and pharmacists. The SAC also maintains an in-house computerized database of all HIV-infected patients, established in 1989, containing relevant all patient's demographic, clinical, and treatment data [18,22–24]. This study used a longitudinal cross-sectional design.

### Patient clinical and demographic variables

Multiple variables were extracted from the SAC database (Table 1) that included: sex, age, continent of birth, years of education, current employment status, sexual orientation, estimated duration of HIV-1 infection (derived from

date of first HIV-1 seropositive test), presence or absence of AIDS, current and nadir CD4<sup>+</sup> T-cell counts, current/peak plasma viral load, current, and past ART (including ddI, ddC, and d4T exposure), polypharmacy (≥5 non-antiretroviral drugs) neurocognitive performance as assessed by neuropsychological testing including the presence or absence of HIV-associated neurocognitive disorders (HAND); co-morbidities (e.g. cardiovascular disease, hepatitis C virus seropositivity); past and present substance use (e.g. alcohol, marijuana, cocaine, heroin, methamphetamine, and other illicit substances); medical conditions: diabetes (types 1 and 2), cardiovascular disease

**Table 1. Clinical, laboratory, and sociodemographic variables for the non-neuropathy, distal sensory polyneuropathy, mononeuropathy, and all peripheral neuropathies groups.**

Variable <sup>a,b</sup>	NNP, <i>n</i> = 408	DSP, <i>n</i> = 90	MNP, <i>n</i> = 21	PNP, <i>n</i> = 111	<i>P</i> value <sup>c</sup>
Age (years)	46.79 (11.4)	52.73 (9.38)	49.91 (10.76)	52.19 (9.67)	<i>a</i> <sup>d</sup>
Sex (female)	54 (13.24%)	12 (13.33%)	9 (42.86%)	21 (18.92%)	<i>b</i>
Birth Continent (North America)	310 (75.98%)	72 (80%)	16 (76.19%)	88 (79.28)	
Height (cm)	174.18 (9.02)	174.05 (8.02)	171.19 (9.4)	173.48 (8.34)	
BMI	26.59 (4.88)	26.62 (6.69)	27.3 (5.25)	26.76 (6.41)	
Education (years)	13.87 (2.65)	13.03 (3.01)	13.76 (2.32)	13.17 (2.9)	<i>a</i>
Employed	0.7 (0.56)	0.5 (0.5)	0.62 (0.5)	0.52 (0.5)	<i>a</i>
Employment (h/week)	27.25 (22.36)	20.16 (23.13)	22.6 (20.9)	20.62 (22.65)	<i>a</i>
Sleep (h/night)	6.81 (1.64)	6.29 (1.83)	6.21 (1.3)	6.28 (1.73)	<i>a</i>
Cigarette use	88 (21.57%)	25 (27.78%)	7 (33.33%)	32 (28.83%)	
Substance use	315 (77.21%)	63 (70%)	16 (76.19%)	79 (71.17%)	
Peak viral load (log <sub>10</sub> copies/ml)	4.62 (1.05)	4.96 (1.15)	4.63 (1.22)	4.89 (1.17)	<i>a</i>
Current viral load (log <sub>10</sub> copies/ml)	1.85 (0.71)	2.09 (1.07)	1.76 (0.58)	2.03 (1.00)	
Nadir CD4 <sup>+</sup> T cells (10 <sup>9</sup> /L)	0.22 (0.16)	0.18 (0.16)	0.2 (0.13)	0.18 (0.15)	<i>a</i>
Current CD4 <sup>+</sup> T cells (10 <sup>9</sup> /L)	0.56 (0.25)	0.62 (0.29)	0.63 (0.42)	0.62 (0.32)	
AIDS-defined	195 (48.87%)	53 (61.63%)	11 (52.38%)	64 (59.81%)	<i>a</i>
HIV-1 duration (years)	10.15 (7.79)	17.19 (8.26)	14.22 (6.98)	16.65 (8.09)	<i>a</i>
ART use	384 (94.12%)	88 (97.78%)	21 (100%)	109 (98.2%)	
HAND	71 (19.03%)	26 (29.89%)	7 (35%)	33 (30.84%)	<i>a</i>
Neuropathic pain	102 (25%)	63 (70%)	19 (90.48%)	82 (73.87%)	<i>a, b</i>
HQoL	3.63 (1.03)	3.19 (1.03)	3.43 (0.81)	3.23 (0.99)	<i>a</i>
NPZ Score	-0.39 (0.73)	-0.58 (0.8)	-0.61 (0.57)	-0.58 (0.76)	<i>a</i>
PHQ-9	6.76 (6.37)	8.92 (6.43)	8.95 (5.77)	8.93 (6.28)	<i>a</i>
Polypharmacy	2.7 (4.32)	6.54 (6.01)	3.9 (4.94)	6.05 (5.9)	<i>a</i>
HbA1c (%)	5.57 (1.01)	6.06 (1.24)	5.58 (0.51)	5.96 (1.14)	<i>a</i>
Diabetes	23 (5.64%)	26 (28.89%)	3 (14.29%)	29 (26.13%)	<i>a</i>
Insulin use (year)	9.08 (117.74)	10.96 (52.65)	0 (0)	8.88 (47.56)	<i>a</i>
CVD	96 (23.53%)	38 (42.22%)	8 (38.1%)	46 (41.44%)	<i>a</i>
Lipodystrophy	29 (7.11%)	15 (16.67%)	4 (19.05%)	19 (17.12%)	<i>a</i>
Dyslipidemia	98 (24.02%)	42 (46.67%)	8 (38.1%)	50 (45.05%)	<i>a</i>
Syphilis seropositivity	80 (19.61%)	8 (8.89%)	1 (4.76%)	9 (8.11%)	<i>a</i>
Malignancy	26 (6.37%)	9 (10%)	1 (4.76%)	10 (9.01%)	
Vitamin B12	410.9 (213.82)	464.31 (293.83)	324.38 (145.99)	435.89 (275.32)	
d4T (days)	148.12 (512.65)	493.33 (834.61)	302.33 (606.11)	457.2 (797.51)	<i>a</i>
ddC (days)	18.24 (111.19)	49.47 (154.84)	79.48 (311.69)	55.14 (192.88)	<i>a</i>
ddl (days)	74.98 (381.9)	390.41 (938)	49.62 (221.05)	325.94 (859.5)	<i>a</i>
d4T/ddC/ddl	64 (15.69%)	42 (46.67%)	7 (33.33%)	49 (44.14%)	<i>a</i>
Pregabalin (years)	6.76 (55.15)	18.17 (87.81)	0 (0)	14.73 (79.31)	<i>a</i>
Lithium (years)	108.72 (2117.58)	12.68 (75.39)	0 (0)	10.28 (68)	<i>a</i>

ART, antiretroviral therapy; CNS, central nervous system; d4T, stavudine; ddC, zalcitabine; ddl, didanosine; DSP, distal sensory polyneuropathy; HbA1c, hemoglobin A1c; HIV-1, HIV type 1; MNP, mononeuropathy; NNP, non-neuropathy; PNP, peripheral neuropathy; TSH, thyroid-stimulating hormone.

<sup>a</sup>Data in parentheses indicate mean and SD (continuous variables) or occurrence and percentages (categorical variables).

<sup>b</sup>BMI, CPE, CNS penetration effectiveness; CVD, cardiovascular disease; HAND, HIV-associated neurocognitive disorders; HQoL, health quality of life assessment; NPZ, neuropsychological z-score; PHQ-9, patient health questionnaire.

<sup>c</sup>NS, non-significant. Other statistically non-significant variables included sexual orientation (heterosexual versus bi-/homosexual), CPE ranking, toxoplasmosis seropositivity, documented seroconversion illness, TSH and folate levels, as well as medications' durations (capsaicin, nitroglycerin, isosorbide, ritonavir, darunavir, atazanavir, Kaletra, metronidazole, vincristine, vinblastine); none displayed significant differences between groups at *P* < 0.05.

<sup>d</sup>Univariate tests were conducted using Mann-Whitney *U* test for continuous data and Fisher's exact test for categorical data. *a* and *b* refer to the significant variables (*P* < 0.05) when comparing NNP versus PNP groups, and DSP versus MNP groups, respectively.

(hypertension, heart failure, myocardial infarction), hypothyroidism, dyslipidemia, lipodystrophy, malignancy, syphilis, and toxoplasma serology. Data on all prescription and over-the-counter medications in addition to ART (but excluding nutritional supplements) were collected. Location of birth was classified by continent.

### Statistical analyses

Demographic and clinical comparisons between groups were performed using univariate and multivariate methods, as well as a principal component analysis (PCA). Univariate tests were conducted using Mann–Whitney *U* test for continuous data and Fisher's exact test for categorical data. For multivariate methods, a logistic regression model was first applied using all demographic and clinical variables to predict neuropathy status of patients. We applied the synthetic minority oversampling technique (SMOTE) [25] to ensure balanced datasets. The SMOTE algorithm was applied only to the training set, and the turning parameters in each classifier were selected using five-folded cross validation. To assess the limited predictive performance of logistic regression, an exploratory PCA was implemented to assess the possibility of linear separation of non-neuropathy (NNP) and PNP groups, leading to the construction of a classifier to differentiate NNP and PNP groups. Thus, we sought a more informative statistical approach by applying various machine learning classification algorithms. Multiple classifiers were implemented using the package 'mlr' [21] in the R project for statistical computing (version 3.6.3) ([www.r-project.org](http://www.r-project.org)) including for univariate analysis and multiple linear regression. A random forest model was adopted because of its balance between robust prediction performance and straightforward interpretability. Mean decreases in accuracy (MDA) were computed to measure the importance level of each variable in the random forest model between the PNP versus NNP groups. Partial dependence plots were computed to visualize the marginal effect of each variable on PNP.

### Data availability

All data presented within the present article are available with accompanying accession numbers upon request to qualified investigators for secondary analyses.

## Results

### Study groups

Of the 519 study patients, 21.4% were diagnosed with PNP (Table 1). This group includes both those with predominantly DSP ( $n=90$ ) and those with MNP ( $n=21$ ). The remaining 408 patients had neither signs nor symptoms of a PNP. Clinical, laboratory and demographic variables were compared for the MNP

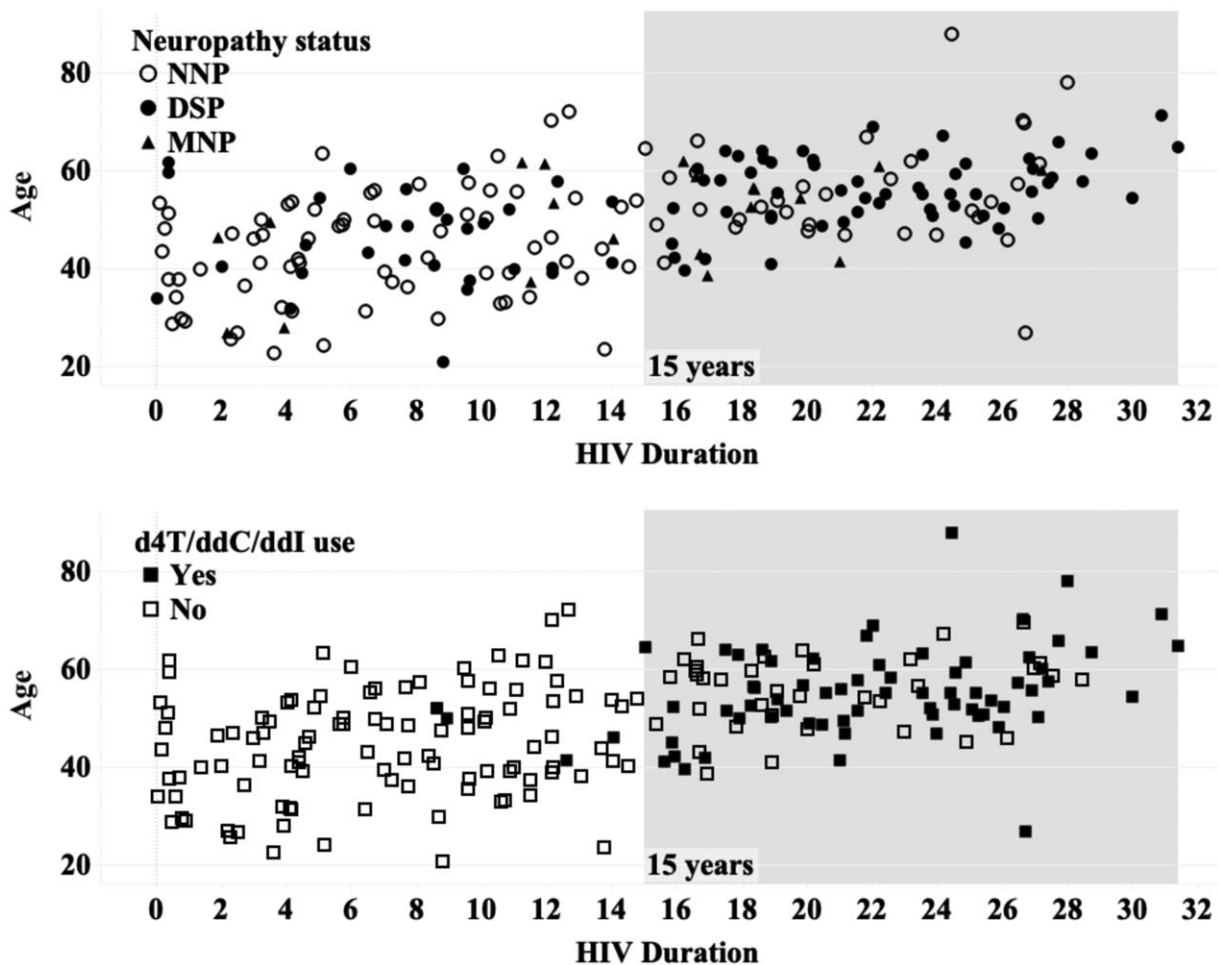
and DSP groups. Significant differences in sex and neuropathic pain frequency were detected with more females and reduced neuropathic pain reported in the MNP group (Table 1). To increase sample sizes for higher statistical power, the MNP and DSP groups were grouped together and the PNP versus NNP groups were compared (Table 1). The univariate comparisons among NNP and PNP groups revealed that 28 of all examined variables ( $n=70$ ) differed significantly including age, HIV-1 duration, d4T/ddC/ddI exposure, diabetes, substance use, dyslipidemia, quality of life, etc. (Table 1). Thus, the NNP and PNP groups were highly differentiated phenotypically when compared using univariate analyses. To explore further the different patterns of PNP among patients with shorter ( $\leq 15$  years) and longer ( $> 15$  years) estimated HIV-1 duration because of ageing effects as well as practice changes in ART drug use, we conducted univariate analyses on these two groups separately. Scatterplots comparing age and HIV-1 duration by neuropathy status (NNP, DSP, or MNP) or d4T/ddC/ddI exposure revealed trends of increased frequencies in DSP with duration and age (Fig. 1a) and a greater exposure to neurotoxic ARTs with age and duration (Fig. 1b). Indeed, while neuropathic pain, polypharmacy, and duration of pregabalin exposure were shared variables for both epochs, multiple other variables differed for PNP among patients with short versus long estimated HIV-1 duration (Table 2). These latter analyses highlighted the impact of both age and duration HIV-1 seropositivity on the development of neuropathy in this intensively treated cohort.

### Principal component analysis

In view of the complexity of multiple variables and to compare the clinical groups, we used a multi-dimensional scaling technique based on PCA to explore and visualize the pattern of association and the potential for building a multivariate predictive model for PNP. To ensure the PCA was straightforward and exploratory, we focused on continuous variables revealing the first two components accounted for 23.45% of the variance (Fig. 2). The PNP group had higher scores on component one, characterizing patients who were older with a longer durations since diagnosis of HIV-1 infection, had higher peak viral loads and a lower nadir CD4<sup>+</sup> T-cell levels, worked less, reduced sleep time, lower HQoL scores and greater past exposure to neurotoxic ART (d4T, ddC, or ddI). A loading plot was constructed that displayed the direction vectors in the PCA (Supplementary Fig. 1, <http://links.lww.com/QAD/C153>). Thus, patterns in the PCA plot showed the potential for building a classifier for PNP using multiple clinical and demographic variables.

### Multivariate analyses

Due to the imbalance of subjects within the present dataset, NNP ( $n=4$ ); PNP ( $n=1$ ), we applied the SMOTE to balance the dataset. We compared a broad range of classifiers, which yielded differential



**Fig. 1.** Scatterplot of age and HIV-1 duration of infection by neuropathy status (non-neuropathy, distal sensory polyneuropathy, and mononeuropathy) (a) or didanosine/zalcitabine/stavudine exposure. (b) Each point represents a single patient with different shapes (non-neuropathy, ○; distal sensory polyneuropathy, ●; mononeuropathy, ▲; didanosine/zalcitabine/stavudine exposure, □; didanosine/zalcitabine/stavudine non-exposure, ■). The shadowed areas represent patients with HIV-1 infection duration greater than 15 years.

performances for several classifiers that were suitable for our dataset. Four evaluation metrics [accuracy, area under the receiver operating characteristics curve (AUROC)], true positive (TPR), and true negative (TNR) rates were derived and were computed on the testing set (25% of the original data) (Table 3). The AUROC (Fig. 3a) was more informative than the accuracy measures because of the imbalance in the dataset; the TPR and TNR measured the relative classification accuracy for patients in the PNP and NNP groups, respectively.

Of importance, all classifiers reached an accuracy of greater than 70% and an AUROC of greater than 77%; the performance of classifiers generally improved with an increase in the complexity. The linear classifier (logistic regression) was outperformed by the advanced and informative ensemble-based classifiers (random forest, adaptive boost); the TPR was lower than TNR for all classifiers, implying that it was more difficult to classify

PNP patients. Random forest, logit boost, and adaptive boost all displayed efficient performances (with TPR and TNR values >70%, and AUROC >80%). In addition to classification performances, the interpretability of a classifier was of equal importance. We focused specifically on two areas: the contribution of each variable to the classification process and the complexity of the classifier. Based on these two criteria, two classifiers were selected for further discussion including logistic regression and random forest (Table 3). Although logit boost and adaptive boost had better numerical performances on this test set, the random forest analyses were presented herein because of the intuitive and interpretable tree-based structure (Supplementary Fig. 2, <http://links.lww.com/QAD/C154>).

### Logistic regression

The multiple logistic regression exhibited a classification accuracy of 73.5% while the TPR was 63.6%. Several

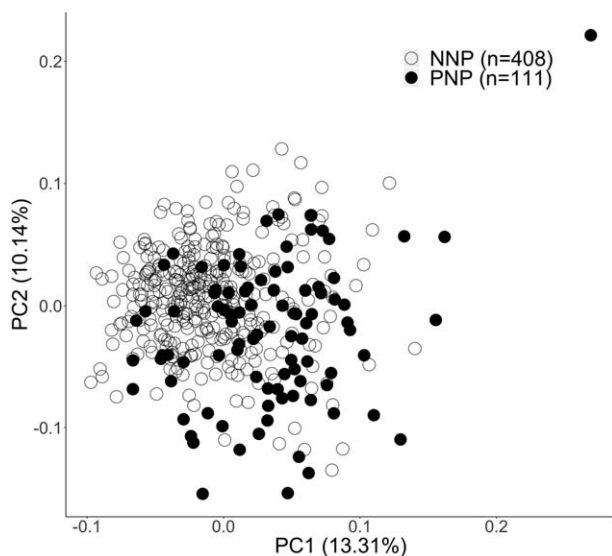
**Table 2. Clinical and demographic variables showing significance for developing peripheral neuropathy based on HIV type 1 duration (< or >15 years) ( $P < 0.05$ )\*.**

HIV-1 duration ( $\leq 15$ years)	HIV-1 duration ( $> 15$ years)
<b>Neuropathic pain (4.36, 35.63)</b>	Diabetes (2.06, 11.86)
Quality of life (-0.97, -0.33)	<b>Neuropathic pain (1.56, 7.75)</b>
<b>Polypharmacy (1.26, 5.19)</b>	<b>Polypharmacy (0.42, 3.86)</b>
PHQ-9 (-10, -6.74)	Viral load (-3.772, 4.38)
AIDS (1.44, 7.64)	<b>Pregabalin duration (-6.66, 41.22)</b>
<b>Pregabalin duration (-11.30, 10.31)</b>	Syphilis seropositivity (1.02, 10.77)
NPZ score (-0.65, -0.12)	d4T duration (-62.99, 484.95)
Nadir CD4 <sup>+</sup> T cell (-0.02, -0.13)	Cardiovascular disease (1.05, 3.62)
Nitroglycerin duration (-4.78, 14.13)	
Vincristine duration (-2.71, 8.01)	
Peak viral load (log <sub>10</sub> copies/ml) (-5.03, 5.99)	
Cigarette use (1.19, 5.51)	
Sleep (h/night) (-1.15, -0.08)	
Education (year) (-2.14, -0.07)	

d4T, stavudine; HIV-1, HIV type 1; NPZ score, neuropsychological z-score; PHQ-9, patient health questionnaire.

\*The number in the brackets is the confidence interval for the mean (for continuous variables, based on the Student *t* test) or the odds ratio (for binary variables). The variables that are significant for both columns are in **bold**.

variables differed significantly between groups in this analysis, and the PNP group was associated with neuropathic pain, higher viral load, higher peak viral load, higher nadir CD4<sup>+</sup>, diabetes, and syphilis, past



**Fig. 2. Principal component analyses: non-neuropathy patients (non-neuropathy,  $n = 408$ ) and peripheral neuropathy ( $n = 111$ ) groups.** The plots show the first two principal components, which accounted for about 20% of variation within all patients. Each point represents a single patient with a specific circle size, shape, and filling that indicate neuropathy status (non-neuropathy, ○) and (peripheral neuropathy, ●).

**Table 3. Comparison of performances for different classifiers.**

Classifier <sup>a</sup>	Accuracy (%) <sup>b</sup>	AUROC (%) <sup>c</sup>	True positive rate (%) <sup>d</sup>	True negative rate (%) <sup>e</sup>
Logistic regression	73.5	77.1	63.6	76.2
kNN <sup>f</sup>	78.7	78.9	57.6	84.4
Naive Bayes	80.6	82.1	42.4	<b>91.0</b>
Bayes net	78.7	84.0	57.6	84.4
Random forest	78.71	83.2	70.3	81.1
Logit boost	<b>82.58</b>	86.6	<b>81.8</b>	82.8
Adaptive boost	81.29	<b>87.4</b>	66.7	85.2

<sup>a</sup>All evaluation metrics were computed on the testing set (25% of the original data). The SMOTE algorithm was applied only to the training set, and the tuning parameters in each classifier were selected using five-fold cross validation. The highest number in each column were boldfaced.

<sup>b</sup>Accuracy: the percentage of correct assessments.

<sup>c</sup>AUROC: the area under the receiver operating characteristics curve measures the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.

<sup>d</sup>True positive rate: the percentage of correct assessments for patients with neuropathy.

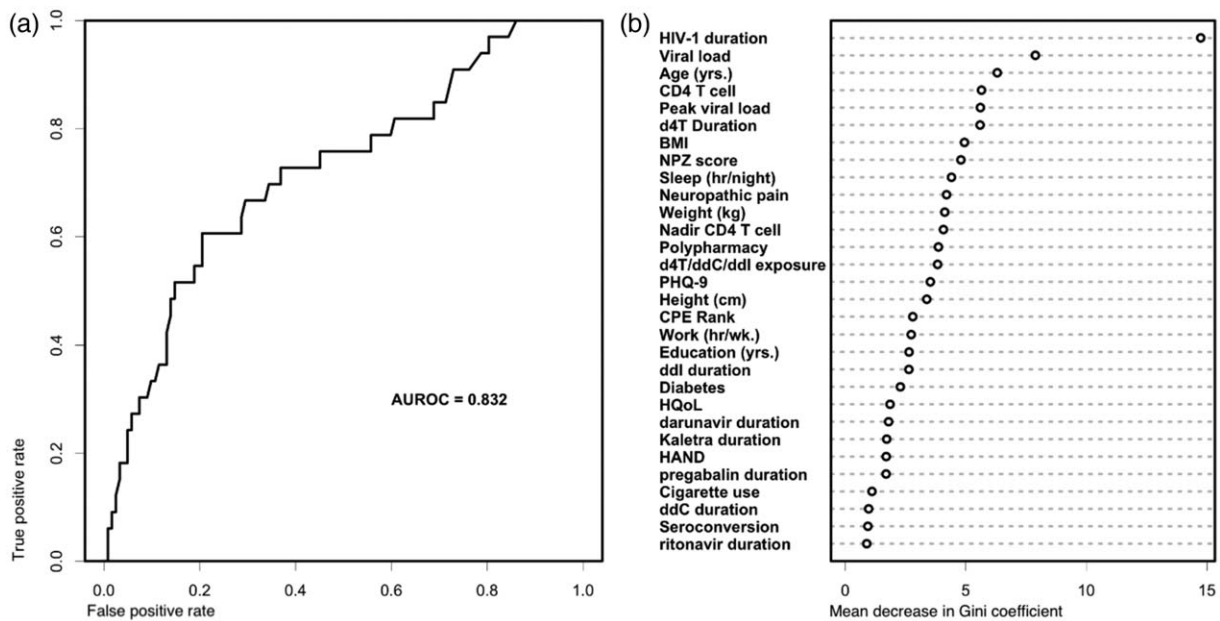
<sup>e</sup>True negative rate: the percentage of correct assessments for patients without neuropathy.

<sup>f</sup>kNN, *K*-nearest neighbor.

d4T/ddC/ddI exposure, longer ddI duration, longer HIV-1 d duration, and dyslipidemia.

### Random forest

The random forest (RF) is an ensemble-based classifier that operates by constructing a multitude of decision trees and outputting the result based on the majority votes of all the decision trees. A decision tree is a graphic representation of all possible outcomes to a decision based on given conditions although in the present analyses, it does not represent a diagnostic algorithm. A random forest classifier was composed of multiple randomly generated similar trees. To explore which variables were more important in the RF models, the MDA values was computed for each variable. The MDA reflected the relative loss of accuracy by the random permutation of one variable. The absolute values of MDA values assign the variables based on relative importance in differentiating the PNP versus NNP groups. Duration since HIV-1 diagnosis, viral load, age, current CD4<sup>+</sup> T-cell count, peak viral load, BMI, neuropsychological z-score (NPZ) score, sleep, and the presence of neuropathic pain were among the top variables identified in the relative importance plot (Fig. 3b). Thus, the MDA comparisons yielded a different profile of predictive variables from the logistic regression analyses offering a distinct perspective on the variables contributing to the development of PNP. Partial dependence plots were also presented (Supplementary Fig. 3A–D, ordered by MDA value, <http://links.lww.com/QAD/C155>) to illustrate the effect of each variable on predicting neuropathy in the random forest analysis.



**Fig. 3. Area under receiver operating characteristic curve and relative variable importance plot.** (a) non-neuropathy versus peripheral neuropathy receiver operating characteristics curve showed an area under the receiver operating characteristic curve value = 0.832. (b) The relative variable importance plots for non-neuropathy versus peripheral neuropathy groups depicts the mean decreases in accuracy values for both continuous and categorical variables that contributed to the differences between the non-neuropathy versus peripheral neuropathy groups.

## Discussion

We undertook the first in depth analysis of PNP in terms of type and predictive variables among patients with HIV-1 infection receiving ART using contemporary machine learning tools that were compared to conventional statistical approaches. The application of machine learning tools not only resulted in better classification performances (Table 3) but also led to the discovery of both known and new variables as principal variables for developing PNP including duration of HIV-1 infection, viral load and CD4<sup>+</sup> T-cell levels at the time of assessment, age as well as duration of past d4T exposure. Indeed, the prevalence of PNP in this cohort was low at 21.4% and was mainly evident as DSP (18%). The importance of estimated HIV-1 duration in the random forest analysis (Fig. 3b) led us to uncover differences in predictive variables for patients with HIV-1 infection for more or less than 15 years reflecting the evolution in ART use together with an ageing population.

The current study builds on to several recent studies of PNP among PWHs in different settings within the USA [26], India [27], West Africa [28], and an international multi-site study [29]. Remarkably, the prevalence rates of PNP were similar across the different studies regardless of the study location and the use of ART despite a wide range of clinical tools to diagnosing PNP [30]. Predictive variables for these studies, which largely focused on DSP were similar to earlier studies and included age, CD4<sup>+</sup> T-

cell nadir and duration of infection. These studies also emphasized the substantial adverse impact of PNP on both employment and quality of life [31], findings that we also documented. However, the effects of prior exposure to neurotoxic nucleoside ART medications was not apparent in some of these studies despite a clear effect in our study. This difference might be due to limited availability of detailed historical information on all past ART regimens or less likely due to different clinical management patterns. Regardless of the explanation, our findings highlight a legacy effect of previous ART exposure and the presence of PNP.

In the current era of increased volume and complexity of data available in medicine, machine learning tools have attracted more attention due to their capacity to delineate non-intuitive patterns in large datasets and apply these findings to tasks such as diagnosis and clinical management [32]. When using machine learning tools, besides achieving superior numerical performances, it is equally (or sometimes more) important to discover useful and interpretable patterns in datasets that strengthen the understanding of a disorder and facilitate clinical decision-making. In this study, we analyzed data using both conventional methods (univariate tests, logistic regression) and machine learning tools. While the univariate analysis identified variables associated with PNP (Table 1), the machine learning tools offered not only classification algorithms for PNP with AUROC values as high as 87.4% (Table 3) but also more insights on

the determinants of PNP among PWH. The random forest classifier operates by constructing a multitude of decision trees. To visualize what one of the trees might look like, a tree was built using 4 selected variables (estimated HIV-1 duration, diabetes, neuropathic pain and peak viral load) (Supplementary Fig. 2, <http://links.lww.com/QAD/C154>). The percentages (in the box) show the classification accuracy, which ranges from 54.16 to 89.6%. It is warrants mentioning that the tree is not intended to be an algorithm for diagnostic decision-making and does not have immediate clinical application. However, from this simple decision tree, it is apparent that PNP was related to longer HIV-1 duration, higher peak viral loads and diabetes. Notably, the prevalence of PNP was 40% among PWH with HIV-1 duration longer or equal to 15 years and only 11% for patients when HIV-1 duration was less than 15 years (Fig. 2a). Furthermore, a longer duration of HIV-1 infection was associated with more frequent d4T/ddC/ddI exposure (Fig. 2b). Different predictive variable profiles for PNP were evident among patients with shorter ( $\leq 15$  years) versus longer ( $> 15$  years) documented durations of HIV-1 infected. Indeed, the predictive variables differed for neuropathy for patients with short versus long HIV-1 duration (Table 2). The shared variables included neuropathic pain, polypharmacy ( $\geq 5$  non-ART medications, Supplementary Fig. 4, <http://links.lww.com/QAD/C156>) and pregabalin duration. For the cohort with shorter documented duration of HIV-1 infection, there were more variables related to mental health measurements such as HQoL, PHQ-9, and NPZ score, as well as variables including cigarette use, sleep and education. For the cohort with longer HIV-1 duration, most variables were linked directly to comorbidities such as diabetes, syphilis and cardiovascular disease. The differing patterns distinguishing these two groups highlighted the potential of machine learning tools in extending the knowledge of neurological disorders and enabling informed clinical decision-making.

The current study faced several limitations. While assessing prevalence of PNP in a general population of HIV-1 infected patients in active care, the actual number of patients with PNP was low and complicated by concurrent co-morbidities with overlapping effects such as diabetes. By pooling patients with DSP and MNP as PNP, distinguishing predictive variables might have been overlooked. Furthermore, the imbalance in group sizes, NNP ( $n = 408$ ) versus PNP ( $n = 111$ ), required statistical manipulations (e.g. SMOTE) to permit comparisons of variables using the machine learning tools herein, which could be misleading. The current study did not use a formal neuropathic pain scale and relied on clinically relevant signs and symptoms, which precluded comparison with prior studies of neuropathic pain. Finally, while longitudinal data were available for this study, it was not truly prospective, given that serial examinations were not performed, which raises the possibility of overlooking

converging factors that influence the development of PNP.

As PNP remains a major comorbidity among people with HIV/AIDS in high-income, middle-income, and low-income countries, it is imperative to understand both its determinants and outcomes, especially the impact of neuropathic pain. The current global opiate use crisis amplifies this priority [33] because opiate use can complicate the already complex care of people with PNP and neuropathic pain. Indeed, a deeper understanding of both pathogenesis and clinical factors defining PNP is required for addressing this issue. Future studies involving prospective and rigorous clinical assessments of PNP coupled with molecular analyses would enable a deeper appreciation of the predictive variables and potential diagnostic and/or therapeutic approaches.

## Acknowledgements

A Canadian Institutes of Health Research Team Grant, Canadian HIV-Ageing Multidisciplinary Programmatic Strategy (CHAMPS) in NeuroHIV (E.F., M.J.G., and C.P.), supported these studies. C.P. and L.K. were supported by Canada Research Chairs in Neurological Infection & Immunity and Statistical Learning, respectively. The authors thank the patients and staff at the Southern Alberta Clinic for their willingness to participate in the study.

## Conflicts of interest

All authors report that they have no financial or personal interests related to the contents of the present article.

## References

1. Kaku M, Simpson DM. **Neuromuscular complications of HIV infection.** *Handb Clin Neurol* 2018; **152**:201–212.
2. Nath A. **Neurologic complications of human immunodeficiency virus infection.** *Continuum (Minneapolis)* 2015; **21**:1557–1576.
3. Stavros K, Simpson DM. **Understanding the etiology and management of HIV-associated peripheral neuropathy.** *Curr HIV/AIDS Rep* 2014; **11**:195–201.
4. Roda RH, Hoke A. **Mitochondrial dysfunction in HIV-induced peripheral neuropathy.** *Int Rev Neurobiol* 2019; **145**:67–82.
5. Asahchop EL, Branton WG, Krishnan A, Chen PA, Yang D, Kong L, et al. **HIV-associated sensory polyneuropathy and neuronal injury are associated with miRNA-455-3p induction.** *JCI Insight* 2018; **3**:e122450.
6. Acharjee S, Noorbakhsh F, Stemkowski PL, Olechowski C, Cohen EA, Ballanyi K, et al. **HIV-1 viral protein R causes peripheral nervous system injury associated with in vivo neuropathic pain.** *FASEB J* 2010; **24**:4343–4353.
7. Mangus LM, Weinberg RL, Knight AC, Queen SE, Adams RJ, Mankowski JL. **SIV-induced immune activation and metabolic alterations in the dorsal root ganglia during acute infection.** *J Neurobiol Exp Neurol* 2019; **78**:78–87.
8. Power C. **Neurologic disease in feline immunodeficiency virus infection: disease mechanisms and therapeutic interventions for NeuroAIDS.** *J Neurovirol* 2018; **24**:220–228.



9. Saag MS, Benson CA, Gandhi RT, Hoy JF, Landovitz RJ, Murgavero MJ, et al. **Antiretroviral drugs for treatment and prevention of HIV infection in adults: 2018 recommendations of the International Antiviral Society-USA Panel.** *JAMA* 2018; **320**:379–396.
10. Gabuzda D, Jamieson BD, Collman RG, Lederman MM, Burdo TH, Deeks SG, et al. **Pathogenesis of aging and age-related comorbidities in people with HIV: highlights from the HIV ACTION workshop.** *Pathog Immun* 2020; **5**:143–174.
11. Karris MY, Berko J, Mazonson PD, Loo TM, Spinelli F, Zolopa A. **Association of pain and pain medication use with multiple characteristics of older people living with HIV.** *AIDS Res Hum Retroviruses* 2020; **36**:663–669.
12. Asahchop EL, Akinwumi SM, Branton WG, Fujiwara E, Gill MJ, Power C. **Plasma microRNA profiling predicts HIV-associated neurocognitive disorder.** *AIDS* 2016; **30**:2021–2031.
13. Tu W, Chen PA, Koenig N, Gomez D, Fujiwara E, Gill MJ, et al. **Machine learning models reveal neurocognitive impairment type and prevalence are associated with distinct variables in HIV/AIDS.** *J Neurovirol* 2020; **26**:41–51.
14. Tymchuk S, Gomez D, Koenig N, Gill MJ, Fujiwara E, Power C. **Associations between depressive symptomatology and neurocognitive impairment in HIV/AIDS.** *Can J Psychiatry* 2018; **63**:329–336.
15. Gomez D, Power C, Gill MJ, Fujiwara E. **Determinants of risk-taking in HIV-associated neurocognitive disorders.** *Neuropsychology* 2017; **31**:798–810.
16. Brew BJ. **The peripheral nerve complications of human immunodeficiency virus (HIV) infection.** *Muscle Nerve* 2003; **28**:542–552.
17. Kaku M, Simpson DM. **HIV neuropathy.** *Curr Opin HIV AIDS* 2014; **9**:521–526.
18. Pettersen JA, Jones G, Worthington C, Krentz HB, Keppler OT, Hoke A, et al. **Sensory neuropathy in human immunodeficiency virus/acquired immunodeficiency syndrome patients: protease inhibitor-mediated neurotoxicity.** *Ann Neurol* 2006; **59**:816–824.
19. Hurley RW, Adams MC, Benzon HT. **Neuropathic pain: treatment guidelines and updates.** *Curr Opin Anaesthesiol* 2013; **26**:580–587.
20. Baron R, Binder A, Wasner G. **Neuropathic pain: diagnosis, pathophysiological mechanisms, and treatment.** *Lancet Neurol* 2010; **9**:807–819.
21. Stewart JD. *Focal peripheral neuropathies.* 4th ed. West Vancouver, Canada: JBL Publishing; 2010, 50–73.
22. Kim D, Jewison DL, milner GR, Rourke SB, Gill MJ, Power C. **HIV-related neurocognitive impairment in patients receiving highly active antiretroviral therapy, HAART.** *Can J Neurol Sci* 2001; **28**:228–231.
23. Vivithanaporn P, Heo G, Gamble J, Krentz HB, Hoke A, Gill MJ, et al. **Neurologic disease burden in treated HIV/AIDS predicts survival: a population-based study.** *Neurology* 2010; **75**:1150–1158.
24. McCombe JA, Vivithanaporn P, Gill MJ, Power C. **Predictors of symptomatic HIV-associated neurocognitive disorders in universal healthcare.** *HIV Med* 2013; **14**:99–107.
25. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. **SMOTE: synthetic minority over-sampling technique.** *J Artif Intell Res* 2002; **16**:321–357.
26. Ellis RJ, Diaz M, Sacktor N, Marra C, Collier AC, Clifford DB, et al. **Predictors of worsening neuropathy and neuropathic pain after 12 years in people with HIV.** *Ann Clin Transl Neurol* 2020; **7**:1166–1173.
27. Gupta PK, Varun V, Mahto SK, Hansraj, Anand KS, Taneja RS, et al. **Prevalence and predictors of distal symmetric polyneuropathy in patients with HIV/AIDS not on highly active antiretroviral therapy (HAART).** *J Assoc Physicians India* 2020; **68**:23–26.
28. Pupilampu P, Ganu V, Kenu E, Kudzi W, Adjei P, Grize L, et al. **Peripheral neuropathy in patients with human immunodeficiency viral infection at a tertiary hospital in Ghana.** *J Neurovirol* 2019; **25**:464–474.
29. Vecchio AC, Marra CM, Schouten J, Jiang H, Kumwenda J, Supparatpinyo K, et al. **Distal sensory peripheral neuropathy in human immunodeficiency virus type 1-positive individuals before and after antiretroviral therapy initiation in diverse resource-limited settings.** *Clin Infect Dis* 2020; **71**:158–165.
30. Gewandter JS, Gibbons CH, Campagnolo M, Lee J, Chaudari J, Ward N, et al. **Clinician-rated measures for distal symmetrical axonal polyneuropathy: ACTION systematic review.** *Neurology* 2019; **93**:346–360.
31. Yakasai AM, Maharaj SS, Kaka B, Danazumi MS. **Does exercise program of endurance and strength improve health-related quality of life in persons living with HIV-related distal symmetrical polyneuropathy? A randomized controlled trial.** *Qual Life Res* 2020; **29**:2383–2393.
32. Rajkomar A, Dean J, Kohane I. **Machine learning in medicine.** *N Engl J Med* 2019; **380**:1347–1358.
33. Stoicea N, Costa A, Periel L, Uribe A, Weaver T, Bergese SD. **Current perspectives on the opioid crisis in the US healthcare system: a comprehensive literature review.** *Medicine (Baltimore)* 2019; **98**:e15425.