

Original quantitative research

Examining the use of decision trees in population health surveillance research: an application to youth mental health survey data in the COMPASS study

Katelyn Battista, MMath (1); Liqun Diao, PhD (2); Karen A. Patte, PhD (3); Joel A. Dubin, PhD (1,2); Scott T. Leatherdale, PhD (1)

This article has been peer reviewed.

 [Tweet this article](#)

Abstract

Introduction: In population health surveillance research, survey data are commonly analyzed using regression methods; however, these methods have limited ability to examine complex relationships. In contrast, decision tree models are ideally suited for segmenting populations and examining complex interactions among factors, and their use within health research is growing. This article provides a methodological overview of decision trees and their application to youth mental health survey data.

Methods: The performance of two popular decision tree techniques, the classification and regression tree (CART) and conditional inference tree (CTREE) techniques, is compared to traditional linear and logistic regression models through an application to youth mental health outcomes in the COMPASS study. Data were collected from 74 501 students across 136 schools in Canada. Anxiety, depression and psychosocial well-being outcomes were measured along with 23 sociodemographic and health behaviour predictors. Model performance was assessed using measures of prediction accuracy, parsimony and relative variable importance.

Results: Decision tree and regression models consistently identified the same sets of most important predictors for each outcome, indicating a general level of agreement between methods. Tree models had lower prediction accuracy but were more parsimonious and placed greater relative importance on key differentiating factors.

Conclusion: Decision trees provide a means of identifying high-risk subgroups to whom prevention and intervention efforts can be targeted, making them a useful tool to address research questions that cannot be answered by traditional regression methods.

Keywords: *decision trees, population health, survey methods, mental health, youth*

Introduction

Population health surveillance research is often carried out using large-scale survey studies that attempt to assess the impacts of wide-ranging social, economic and environmental factors on various health outcomes. The relationship between these factors and health outcomes is often

characterized by complex interactions that make it impractical to identify any single factor as causal. In the context of youth mental health, outcomes have previously been associated with socioeconomic status,¹ weight status,² dietary behaviours,³ physical activity and sedentary behaviours,⁴ sleep habits,⁵ cannabis use,⁶ bullying,⁷ school connectedness^{8,9} and peer and

family relationships,^{10,11} among other factors. However, most research studies focus on examining the impact of any given factor or domain of factors in isolation; in reality, the underlying interrelationships are likely more complex.

Associations are often examined using regression models, which estimate the association between a predictor and an outcome while controlling for other factors. However, these models are rarely used to estimate complex interactions between factors, due to computational limitations and difficulty in interpretation. Additionally, the resulting model estimates do not

Highlights

- Decision trees can be used within population health research to address important research questions that cannot be answered by traditional regression methods.
- A key advantage of decision trees over regression models is the ability to examine complex interactions among risk factors.
- Decision trees can be used to identify high-risk groups to whom prevention and intervention efforts can be targeted.
- While regression models may have higher prediction accuracy in some settings, decision trees place greater emphasis on key differentiating factors.

Author references:

1. School of Public Health Sciences, University of Waterloo, Waterloo, Ontario, Canada
2. Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada
3. Department of Health Sciences, Brock University, St. Catharines, Ontario, Canada

Correspondence: Katelyn Battista, School of Public Health Sciences, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1; Tel: 519-888-4567 x 46706; Email: kbattista@uwaterloo.ca

allow for the development of risk profiles, that is, separating subjects into subgroups based on certain combinations of risk factors. The identification of high-risk subgroups is important to efficiently target resources and interventions. Decision trees comprise a different class of models that are ideally suited for segmenting populations and examining complex interactions among factors.¹²

Decision trees are commonly used in clinical research that focusses on screening and diagnostics,¹³ with emphasis on prediction. Decision trees are less common in population health research, where the focus is on understanding associations and identifying subgroups for targeting behavioural interventions, though their use is increasing. Within the domain of mental health, recent studies using decision trees have primarily examined associations with depression¹⁴⁻¹⁹ and suicide risk.^{15,20-28}

Two studies examined depression outcomes in youth populations specifically. Hill et al.¹⁶ found that, among students with subclinical depressive symptoms at baseline, friend support was protective against developing major depressive disorder by age 30, with anxiety disorder and substance use disorder increasing risk among those without friend support. Seeley, Stice and Rohde¹⁸ found poor school functioning to be a primary risk factor for major depressive disorder onset among girls with elevated depressive symptoms at baseline, with parental support acting as a protective factor only among girls with low levels of baseline depressive symptoms. Three studies examined suicide ideation in youth populations and found that mediating factors such as family relationships^{22,26} and social support^{22,24} were only protective among students that did not have high levels of depression.

Among the studies mentioned above, few included direct performance comparisons between tree and regression methods. Smaller studies by Burke et al.,²¹ Mitsui et al.¹⁵ and Handley et al.²⁷ found regression models had higher predictive accuracy than corresponding tree models; however, these studies had small sample sizes (ranging from 359 to 2194 participants). Conversely, two larger studies—one by Dykxhoorn et al.²³ examining a longitudinal sample of 11 088 children, and another

by Batterham et al.¹⁷ examining a longitudinal study of 6605 adults—found decision trees to outperform corresponding logistic regression in terms of sensitivity and overall predictive accuracy. Thus, while there is some evidence to suggest that decision trees may have advantages over traditional regression methods in the case of larger sample sizes, there is an overall lack of available evidence within the domain of mental health.

Despite growing use of decision trees, regression models remain commonplace in the population health literature. This results in a missed opportunity to understand the complex interactions among risk factors and to identify high-risk subgroups to which prevention and intervention efforts can be targeted. The aim of this study was therefore to examine the use of decision trees in the analysis of large-scale population health surveillance data. In this paper, we first provide an overview of two popular types of decision tree, the classification and regression tree (CART) and the conditional inference tree (CTREE) techniques. Next, the performance of decision tree models is compared to traditional linear and logistic regression models through an application to youth mental health outcomes in the COMPASS study.²⁹ Tree and regression methods are evaluated based on prediction accuracy and parsimony, with additional considerations given to relative variable importance and model interpretability.

Methods

Background on decision trees

Decision trees are statistical models that examine an outcome of interest by partitioning the sample into distinct subgroups based on predictor variables. The subgroups are determined using a series of binary splits that resemble a tree structure. Various types of decision tree algorithms have been developed;³⁰ this analysis focusses on two popular types of decision tree: CART and CTREE. Methodological overviews of CART and CTREE in the context of epidemiological research have been previously published;^{12,13} a summary of important features follows.

Classification and regression trees

CART is a widely used class of decision tree for categorical (classification) and continuous (regression) outcomes. Originally

developed by Breiman et al.,³¹ CART methods find optimal splits of the sample into subgroups³² such that subjects within a subgroup are similar and subjects across subgroups are as different as possible. Optimal splits are determined by recursively choosing the variables and cut-off levels that produce maximum separation among subgroups and minimal within-group variability with respect to the outcome.³² Continuous and categorical variables may be split multiple times throughout the tree on different cut-points. Splitting occurs until a stopping rule is reached, typically based on minimum subgroup size.^{12,32,33} Through this recursive process, the predictor space is divided into a final set of subgroups, for which the mean outcome value (regression trees) or the percent of the subgroup having the outcome (classification trees) is calculated.³³

A large tree grown by recursively splitting the predictor space tends to overfit the sample data, resulting in poor generalizability. Overfitting is mitigated using tree pruning and a cross-validation procedure, in which the large tree is pruned leading to a sequence of nested subtrees from among which an optimal tree is selected. The most commonly used pruning method is cost complexity pruning, in which an increasing sequence of complexity parameters corresponds to a sequence of nested subtrees with decreasing sizes.^{33,34} The optimal subtree that minimizes the average error based on cross-validation³³ is then chosen. When working with larger samples, the “1-SE” rule is often used to choose the smallest subtree that has an average error within one standard deviation of the overall minimum error.^{12,13}

Conditional inference trees

CTREE is an alternative to CART developed by Hothorn et al.³⁵ While CART chooses the optimal split at each step among all potential variable and splitting points simultaneously, CTREE separates the splitting determination into two steps. First, the optimal variable to split on is chosen based on the strongest association to the outcome. Association to the outcome variable is measured using regression models appropriate for the outcome, for example, linear regression for continuous outcomes and logistic regression for binary outcomes.^{12,35} The covariate with the smallest *p* value is chosen for splitting. Second, the optimal splitting point for that variable is determined.^{12,35} This approach mitigates the selection bias

toward variables with many splitting points often found in CART.^{12,35} This splitting process continues recursively among each subgroup until a stopping rule is reached. As with CART, continuous and categorical variables can be split more than once throughout the tree at different cut-points.

The stopping rule for CTREE is based on a global null hypothesis: the algorithm stops splitting when no covariates have a significant association to the outcome based on a prespecified significance level (α ; α).^{12,35} For large samples, additional stopping criteria based on minimum subgroup sizes can also be used. No pruning is required in CTREE; the global test for significance acts as a means to prevent overfitting.^{12,35}

Application

The relative performance of decision trees and regression methods was compared in the context of population surveillance research using youth mental health data from the COMPASS study.²⁹

Ethics approval, study design and sample

COMPASS is a prospective cohort study designed to collect hierarchical health data from Canadian secondary school students.²⁹ COMPASS has received ethics clearance from the University of Waterloo Research Ethics Board (ORE 30118). Additional details about the COMPASS host study are available in print²⁹ and online (<https://uwaterloo.ca/compass-system>).

We used student-level data from Year 7 (2018/19) of the COMPASS study. The sample consists of 74 501 students from 136 schools in Ontario (61 schools), Alberta (8 schools), British Columbia (15 schools) and Quebec (52 schools). COMPASS uses purposeful sampling to recruit whole-school samples based on their use of active-information, passive-consent parental permission protocols. The participation rate for 2018/19 was 81.9%, with the primary reason for nonparticipation being absenteeism or scheduled spare on the data collection date.

Measures

The COMPASS student questionnaire is a paper-based questionnaire completed by students during class time. The questionnaire is anonymous and self-administered, and students may decline to participate at any time. This study examined 5 mental

health outcome measures related to depression, anxiety and psychosocial well-being (flourishing), as well as 23 core predictor measures related to demographics, body weight, healthy eating, movement behaviours, substance use, bullying, academics and school support, and perceived family and friend support.

Mental health outcomes

Depression

Depression is measured using the Center for Epidemiologic Studies Depression 10-item scale (CESD-10),^{36,37} which has been validated in adolescent populations.³⁸ The CESD-10 is measured as a continuous score ranging from 0 to 30, with higher scores indicating greater degrees of depressive symptomatology and risk of unipolar depression. An additional binary measure of depression is used, with students scoring greater than or equal to 10 classified as having clinically relevant depressive symptoms.

Anxiety

Anxiety is measured using the Generalized Anxiety Disorder 7-item Scale (GAD-7),³⁹ which has been validated in adolescent populations.⁴⁰ The GAD-7 is measured as a continuous score ranging from 0 to 21, with higher scores indicating greater levels of anxiety. An additional binary measure of anxiety is used, with students scoring greater than or equal to 10 classified as having clinically relevant anxiety symptoms.

Flourishing

Flourishing is a component of psychological well-being and is measured using a modified version of Diener's Flourishing Scale (FS),⁴¹ which has been validated in young adults.⁴² The FS is a continuous score ranging from 8 to 40, with higher scores indicating greater levels of flourishing.

Predictor variables

Demographic predictor variables include age, sex, ethnicity and weekly spending money (a proxy for socioeconomic status). Body weight is measured using weight perception and body mass index (BMI) classification. Healthy eating is measured using a binary indicator of whether students eat breakfast daily, as well as the number of servings of fruits and vegetables consumed daily. Movement behaviours are assessed using minutes of average daily moderate-to-vigorous physical activity (MVPA), minutes of total daily screen time and daily minutes of sleep.

Substance use is measured using binary indicators of past-month use of tobacco, e-cigarettes and cannabis, as well as past-month binge drinking. Bullying is measured using two indicators: whether a student was bullied or had bullied others in the past 30 days. Academics and school support are measured using a binary indicator of whether students expect to attend a postsecondary institution, the number of classes skipped in the past four weeks and a continuous school connectedness score (with higher scores indicating higher levels of connection to school). Perceived family and friend support are measured using binary indicators of having a happy home life, feeling able to talk about problems with family and feeling able to talk about problems with friends.

In addition to the student-level measures, additional school-level predictors include total school enrolment, province, school area median income and school urbanicity. Measures of income and urbanicity are taken from Statistics Canada's 2016 census and values linked by school forward sortation area.^{43,44}

Analysis

Individual mental health scale items were person-mean imputed for students missing one or two items. While mean imputation may artificially reduce variance, more complex imputation methods were not used given the primary focus of the analysis on performance rather than inference. Students with missing or outlier values on any variables were removed, resulting in a final complete case sample of 52 350 students. Sample characteristics are provided in Table 1. The sample was randomly split into training (41 795; 80%) and test (10 555; 20%) samples.

CART and CTREE were run for continuous (CESD-10, GAD-7, FS) and binary outcomes (depression, anxiety). CART pruning was performed using 10-fold cross-validation and the 1-SE rule. CTREE significance was set at $\alpha = 0.05$ with a Bonferroni adjustment for multiple testing. Given the large sample size, an additional stopping rule was included for both CART and CTREE to limit the minimum number of observations per bucket to 1% of the sample. Linear and logistic regression models were also run for continuous and binary outcomes including all main effects. Backward variable selection using the Akaike information criterion (AIC) was performed to mimic tree pruning.

TABLE 1
COMPASS Year 7 (2018/19) student sample characteristics

Category	Variable	Levels	n	%	
Total			52 350	100.0%	
Mental health outcomes	CESD-10 scale	Mean (SD)	8.50 (5.85)		
	GAD-7 scale	Mean (SD)	6.02 (5.31)		
	Flourishing scale	Mean (SD)	32.42 (5.39)		
	Depression	No		33 778	64.5%
		Yes		18 572	35.5%
	Anxiety	No		40 568	77.5%
Yes			11 782	22.5%	
Demographic factors	Sex	Female	27 483	52.5%	
		Male	24 867	47.5%	
	Age (years)	12	2 310	4.4%	
		13	4 564	8.7%	
		14	10 282	19.6%	
		15	12 221	23.3%	
		16	12 198	23.3%	
		17	8 628	16.5%	
	Ethnicity	18	2 147	4.1%	
		White	37 370	71.4%	
		Black	1 565	3.0%	
		Asian	5 559	10.6%	
	Spending money	Latin American	1 235	2.4%	
		Other/multi	6 621	12.6%	
		\$0	8 099	15.5%	
		\$1–\$20	12 701	24.3%	
		\$21–\$40	5 796	11.1%	
		\$41–\$100	6 469	12.4%	
	Province	More than \$100	10 067	19.2%	
		Don't know	9 218	17.6%	
Alberta		2 222	4.2%		
British Columbia		7 298	13.9%		
Urbanicity	Ontario	20 450	39.1%		
	Quebec	22 380	42.8%		
	Large urban	28 684	54.8%		
School median income (in thousands CAD)	Medium urban	5 044	9.6%		
	Small urban/rural	18 622	35.6%		
	Mean (SD)	67.33 (17.47)			
School size (in hundreds of students)	Mean (SD)	8.49 (3.52)			

Continued on the following page

TABLE 1 (continued)
COMPASS Year 7 (2018/19) student sample characteristics

Category	Variable	Levels	n	%
Body weight and eating behaviours	Weight perception	Underweight	8 300	15.9%
		About the right weight	31 877	60.9%
		Overweight/obese	12 173	23.3%
	BMI classification	Underweight	985	1.9%
		Normal weight	29 932	57.2%
		Overweight	6 465	12.3%
		Obese	2 843	5.4%
		Not stated	12 125	23.2%
	Eat breakfast daily	No	25 373	48.5%
		Yes	26 977	51.5%
Movement behaviours	Servings of fruits and vegetables	Mean (SD)	2.98 (1.93)	
	Average daily physical activity (min)	Mean (SD)	96.40 (62.14)	
	Screen time (min)	Mean (SD)	350.97 (178.28)	
	Sleep time (min)	Mean (SD)	451.94 (74.78)	
Current substance use	Tobacco use	No	49 349	94.3%
		Yes	3 001	5.7%
	E-cigarette use	No	38 570	73.7%
		Yes	13 780	26.3%
	Binge drinking	No	44 020	84.1%
		Yes	8 330	15.9%
	Cannabis use	No	46 683	89.2%
		Yes	5 667	10.8%
Bullying in the last 30 days	Was bullied	No	46 412	88.7%
		Yes	5 938	11.3%
	Bullied others	No	49 702	94.9%
		Yes	2 648	5.1%
Academics and school support	Expect to attend postsecondary institution	No	12 380	23.6%
		Yes	39 970	76.4%
	Classes skipped in past 4 weeks	0 classes	34 894	66.7%
		1 or 2 classes	10 634	20.3%
		3 to 5 classes	4 246	8.1%
		6 or more classes	2 576	4.9%
	School connectedness score	Mean (SD)	18.67 (3.14)	
Family and peer support	Happy home life	No	10 219	19.5%
		Yes	42 131	80.5%
	Talk about problems with family	No	20 770	39.7%
		Yes	31 580	60.3%
	Talk about problems with friends	No	12 748	24.4%
		Yes	39 602	75.6%

Abbreviations: BMI, body mass index; CAD, Canadian dollars; CESD-10, Center for Epidemiologic Studies Depression 10-item scale; GAD-7, Generalized Anxiety Disorder 7-item scale; min, minutes; SD, standard deviation.

Fitted models from the training sample were applied to the test sample. Predictive performance was compared using adjusted R^2 (R^2_{adj}) and root mean square error (RMSE) for continuous outcomes, and percent classification accuracy (pCA) and area under the receiving operator characteristic curve (AUC) for binary outcomes. R^2_{adj} is the amount of variation explained by the model, adjusted for the number of covariates, such that R^2_{adj} will decrease if inclusion of a given covariate does not substantially increase the explained variation. RMSE is the average of the squared difference between the actual and predicted outcome values.³³ The closer the predicted values are to the true values, the lower the RMSE. pCA simply measures the percentage of observations for which the model correctly assigns the outcome value. AUC (also known as the concordance statistic) is a more sophisticated measure of accuracy that accounts for both the sensitivity and specificity of the model.³² Both measures range from 0 to 1, with higher values indicating higher model accuracy.

Parsimony was evaluated using the number of parameters and unique variables in the model. Relative variable importance measures were calculated based on the decrease in model fit resulting from removing a given variable from each model. For decision trees, this is measured by the sum of the goodness of split for all occurrences where the variable is used as a primary or surrogate split. For linear and logistic regression models, this is measured by the decrease to R^2_{adj} and AUC, respectively.

R version 4.0.3 (R Foundation for Statistical Computing, Vienna, AT) was used for all analyses. The functions “rpart” (package “rpart”) and “ctree” (package “partykit”) were used for CART and CTREE models, respectively. The functions “lm” and “glm” (package “MASS”) were used for linear and logistic regression models, respectively.

Results

The average CESD-10 score in the sample was 8.50 (SD = 5.85) with 33.5% of the sample classified as having clinically relevant depressive symptoms. The average GAD-7 score was 6.02 (SD = 5.31) with 22.5% classified as having clinically relevant anxiety symptoms. The average FS score was 32.42 (SD = 5.39).

Decision tree and regression model comparison

As an illustrative example, the CART and logistic regression model results for the binary anxiety outcome are presented. The final fitted CART tree for the binary anxiety outcome is presented in Figure 1. The model identified 9 final subgroups using 5 unique variables. The primary splitting variable was whether students indicated having a happy home life. Both subgroups were then split based on school connectedness, though different cut-off points were used. Splits were also made for some subgroups on sex, sleep duration and whether the student was bullied. The largest final subgroup was of students who indicated having a happy home life and had school connectedness scores of 17.5 or greater, making up 61% of the sample. Within this group, the probability of having clinically relevant anxiety symptoms was 12.7%, which was the lowest of all groups. The highest-risk subgroup comprised females who indicated not having a happy home life and low school connectedness (< 16.25), with a 64.6% probability of having clinically relevant anxiety symptoms.

The logistic regression model result for anxiety is presented in Table 2. The final model after applying backward variable selection included 20 variables (33 parameters). Like the CART model, having a happy home life (odds ratio [OR]: 0.33; 95% CI: 0.31–0.34), male sex (OR: 0.33; 95% CI: 0.31–0.34) and school connectedness (OR: 0.88; 95% CI: 0.87–0.89) were found to be important predictors. Other factors including minority ethnicity, higher spending money, living in Quebec, small urban or rural urbanicity, “about right” weight perception, eating breakfast daily, higher sleep time and feeling able to talk about problems with family and friends were associated with lower odds of having clinically relevant anxiety symptoms. Older age, eating more fruits and vegetables, higher screen time, current tobacco use and e-cigarette use, being bullied, expecting to attend a postsecondary institution and skipping classes were associated with higher odds of having clinically relevant anxiety symptoms.

Prediction accuracy and parsimony

Prediction accuracy results for continuous outcomes (CESD-10, GAD-7, FS) are presented in Table 3. The linear regression

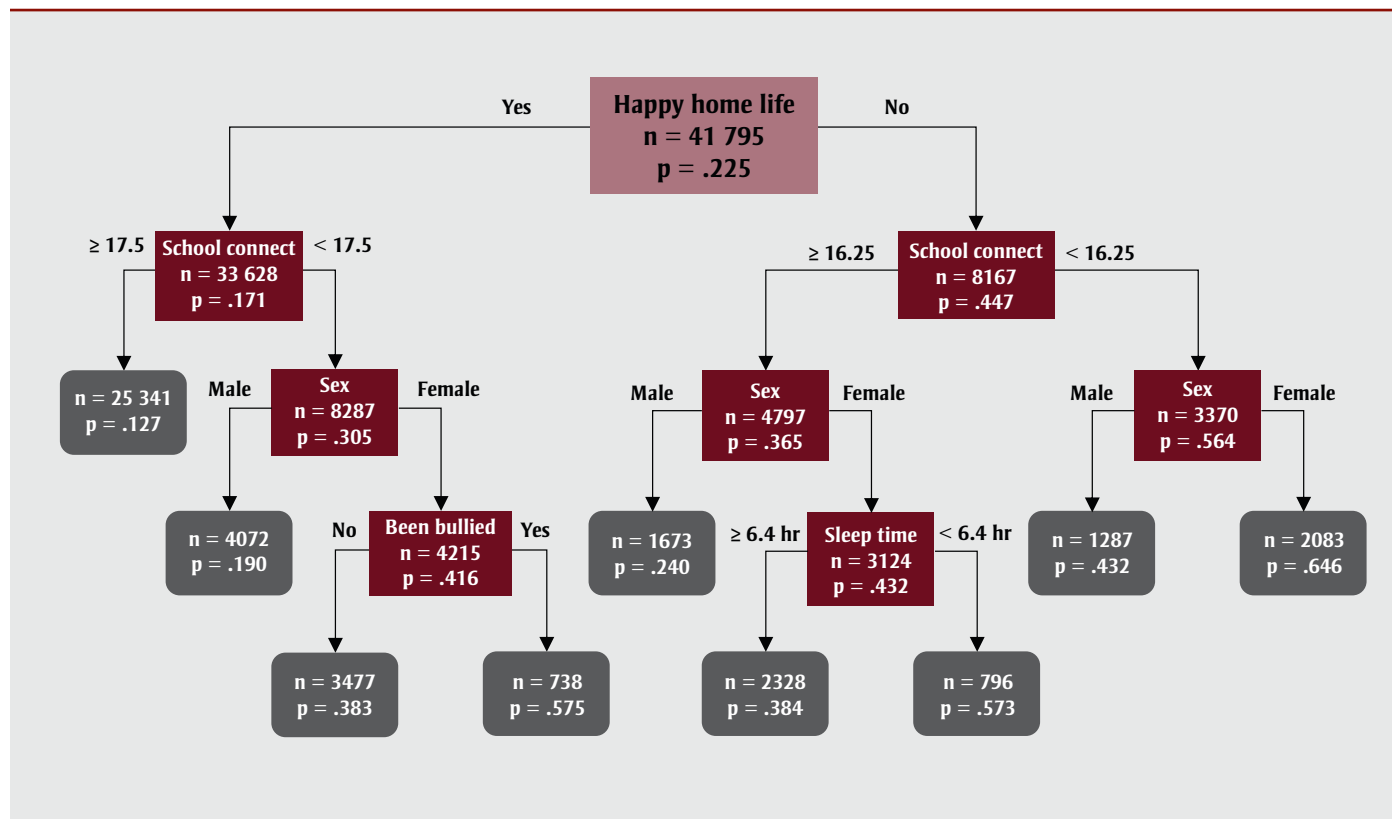
models had the highest test set R^2_{adj} and lowest RMSE for all three outcomes. The R^2_{adj} and RMSE values were similar for CART and CTREE models, with R^2_{adj} consistently 4% to 5% lower than the linear regression results and RMSE 0.13 to 0.19 higher. The CART trees included the fewest unique variables, followed by CTREE, with linear regression models including over twice as many variables. However, the number of final parameters (corresponding to number of splits for tree models) was similar for CART and linear regression, and higher for CTREE models. The absolute value of the R^2_{adj} was relatively low for all models, indicating the predictors explain less than half of the variation in the outcome. Additionally, the R^2_{adj} and RMSE calculated on the test set were similar to the training set for all models, suggesting minimal overfitting.

Prediction accuracy results for binary depression and anxiety outcomes are presented in Table 3. CART produced more parsimonious models than CTREE and logistic regression, using only 9 splits on 6 variables for depression, and 8 splits on 5 variables for anxiety. CTREE produced more complex models, using over 50 splits. The larger difference between number of subgroups and variables used in the CTREE models compared to the CART models is partially due to the model splitting on the same variable multiple times using different cut-points. Logistic regression models included 22 unique variables for depression and 20 for anxiety. Despite the difference in model complexity, the test set pCA and AUC were very similar across models, with logistic regression performing only slightly better. The absolute value of the AUC was 0.71 for depression and ranged from 0.59 to 0.63 for anxiety, which suggests mediocre discriminatory ability. As in the continuous case, training and test set performances were similar, suggesting minimal overfitting.

Relative variable importance

Relative variable importance percentages for continuous outcomes (CESD-10, GAD-7, FS) are presented in Figure 2. For CESD-10 and GAD-7 outcomes, CART, CTREE and logistic regression all consistently identified school connectedness, having a happy home life and sex as the three most important variables. Sleep time also ranked fourth highest in relative importance in all except the anxiety linear regression model, which ranked bullying

FIGURE 1
CART tree for having clinically relevant anxiety symptoms (GAD-7 ≥ 10)



Abbreviations: CART, classification and regression tree; GAD-7, Generalized Anxiety Disorder 7-item Scale; hr, hours; school connect, school connectedness.

Note: n is the number of students in subgroup; p is the percentage of the subgroup with clinically relevant anxiety symptoms.

as fourth highest. However, the CART and CTREE models gave more weight to the highest ranked variables than the linear regression models. CART and CTREE attributed 78% to 87% of the total importance to the top four variables, while linear regression attributed only 47%, with the remainder split more evenly across other variables in the model.

Similar results are seen for FS, though sex is not identified as important in any of the models, while talking about problems with friends is ranked within the top four for all models, family was identified as important for CART and CTREE models, and spending money was identified as important for linear regression. Again, CART and CTREE attributed 86% to 93% of total importance to the top four ranked variables, while linear regression attributed only 43%.

Relative variable importance percentages for binary outcomes are presented in Figure 2. As was seen for continuous outcomes, school connectedness, happy home life and sex were consistently identified as

the three most important variables across depression and anxiety models. Talking about problems with family was ranked as fourth highest for depression across all models, while having been bullied was ranked as fourth highest for all anxiety models. CART attributed 92% to 93% of total importance to the top four variables, while CTREE attributed 79% to 83% and logistic regression attributed 44% to 46%.

Discussion

This study provided a methodological overview and comparison of two types of decision tree, CART and CTREE, to traditional linear and logistic regression methods using a novel application to large-scale youth mental health survey data. This study adds to the limited existing evidence on decision tree performance in this domain^{15,17,21,23,27} by examining a large sample of youth and wide breadth of predictors. This study also examines methodological considerations of decision trees in the context of population surveillance research, in which prediction accuracy must be weighed against model interpretability.

Beyond the subject matter knowledge gleaned from the results of this application to youth mental health, the implications discussed below can be used as a guide for researchers examining other large-scale survey datasets.

In the case of prediction accuracy, for linear scale outcomes linear regression outperformed CART and CTREE, with 4% to 5% higher R^2_{adj} values and 3% to 5% lower RMSE values. The number of model parameters was similar for CART and linear regression, while CTREE resulted in more complex models. However, while CART and linear regression had a similar number of parameters, CART identified far fewer unique variables as significant, with the high number of parameters due to multiple splits on the same continuous predictor variables. In contrast, regression models assumed a linear effect of continuous variables and provided only a single estimate representing the effect of a one-unit increase in the variable, regardless of the starting value.

In the case of binary outcomes, logistic regression models again had higher

TABLE 2
Logistic regression model for odds of having clinically relevant anxiety symptoms (GAD-7 ≥ 10)

Variable	Level	AOR (95% CI)
Sex (ref = female)	Male	0.33 (0.31–0.34)***
Age (years)	per year	1.05 (1.02–1.07)***
Ethnicity (ref = White)	Black	0.5 (0.43–0.59)***
	Asian	0.73 (0.66–0.81)***
	Latin American	0.83 (0.7–0.98)*
	Other/multi	1.01 (0.94–1.09)
Spending money (ref = \$0)	\$1–\$20	0.93 (0.85–1.01)
	\$21–\$40	0.86 (0.77–0.95)**
	\$41–\$100	0.87 (0.79–0.96)**
	More than \$100	0.94 (0.86–1.03)
	Don't know	0.87 (0.79–0.96)**
Province (ref = Alberta)	British Columbia	0.89 (0.77–1.03)
	Ontario	0.92 (0.81–1.05)
	Quebec	0.66 (0.58–0.76)***
Urbanicity (ref = large urban)	Medium urban	1.02 (0.93–1.12)
	Small urban/rural	0.86 (0.80–0.91)***
Weight perception (ref = underweight)	About the right weight	0.78 (0.72–0.84)***
	Overweight	1.03 (0.95–1.12)
Eat breakfast daily	Yes	0.76 (0.72–0.8)***
Servings of fruits and vegetables	per serving	1.03 (1.01–1.04)***
Screen time (hours)	per hour	1.05 (1.05–1.05)***
Sleep time (hours)	per hour	0.83 (0.83–0.83)***
Current tobacco use	Yes	1.12 (1.00–1.25)*
Current e-cigarette use	Yes	1.08 (1.01–1.15)*
Was bullied in last 30 days	Yes	2.03 (1.88–2.18)***
Expect to attend postsecondary institution	Yes	1.16 (1.09–1.24)***
Classes skipped in past 4 weeks (ref = 0 classes)	1–2 classes	1.06 (0.99–1.13)
	3–5 classes	1.16 (1.06–1.28)**
	6 or more classes	1.23 (1.10–1.39)***
School connectedness score	per unit	0.88 (0.87–0.89)***
Happy home life	Yes	0.50 (0.47–0.54)***
Talk about problems with family	Yes	0.73 (0.69–0.77)***
Talk about problems with friends	Yes	0.75 (0.71–0.8)***

Abbreviations: AOR, adjusted odds ratio; CI, confidence interval; ref, reference group.

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

predictive performance than CART and CTREE; however, overall performance was closer than for continuous outcomes, with 1% to 2% higher prediction accuracy and 0% to 3% higher AUC. In these cases, CART produced far more parsimonious models than both CTREE and logistic

regression, both in terms of total parameters and number of unique variables. Previous small studies by Burke et al.,²¹ Mitsui et al.¹⁵ and Handley et al.²⁷ found AUCs ranging 4% to 8% lower for CART than logistic regression, while in contrast, a study by Batterham et. al¹⁷ found AUC

2% higher for CART than logistic regression. While direct comparison of AUC findings from these studies is difficult given the differences in study samples, outcomes and model specifications, it is still noteworthy that across all studies performance between the two techniques did not drastically differ.

Thus, while linear and logistic regression may provide slight advantages in predictive ability, the simpler models generated by CART may be more desirable, particularly for knowledge translation in the context of population health research where the focus is on understanding associations and communicating results to a non-technical audience.

Decision tree and regression models consistently identified the same sets of most important predictors for each outcome, indicating a general level of agreement between methods. However, CART and CTREE weighted the relative importance of these top predictors much higher than the regression models, attributing more than three-quarters of total importance to the top four predictors, compared to regression models, which attributed less than half of total importance to the top predictors. This is in line with the greater parsimony seen in the CART and CTREE models and highlights the ability of decision trees to single out the most important factors.

Additionally, a common limitation of regression models is that factors with high multicollinearity tend to “wash out” when entered simultaneously, leading to inflated variance estimates or variable omission bias, which could cause factors to be overlooked.⁴⁵ This has been seen in past research comparing trees and regression,²³ suggesting that decision tree methods can offer a clearer representation of key factors to aid in decision making. This advantage of parsimony can be particularly beneficial in the domain of population-level disease prevention research, in which a myriad of competing risk factors and confounders may be present.

Higher levels of school connectedness and having a happy home life were consistently identified as key predictors and were associated with lower levels of depression and anxiety and higher flourishing. This is consistent with previous research linking family relationships to

TABLE 3
Prediction accuracy comparison for continuous and binary outcomes for CART, CTREE and regression models

Continuous outcome	Method	# Parameters	# Unique variables	Training R^2_{adj}	Training RMSE	Test R^2_{adj}	Test RMSE
CESD-10	CART	38	9	0.35	4.73	0.33	4.76
	CTREE	57	10	0.36	4.70	0.34	4.73
	Linear reg.	34	20	0.39	4.59	0.38	4.57
GAD-7	CART	39	11	0.28	4.50	0.27	4.55
	CTREE	63	15	0.29	4.49	0.27	4.55
	Linear reg.	40	23	0.32	4.39	0.31	4.42
FS	CART	43	9	0.47	3.94	0.46	3.97
	CTREE	70	12	0.47	3.93	0.46	3.96
	Linear reg.	40	24	0.51	3.79	0.51	3.78
Binary outcome	Method	# Parameters	# Unique variables	Training pCA	Training AUC	Test pCA	Test AUC
Depression	CART	9	6	0.75	0.71	0.74	0.70
	CTREE	53	14	0.75	0.71	0.74	0.70
	Logistic reg.	39	22	0.76	0.71	0.76	0.70
Anxiety	CART	8	5	0.80	0.60	0.79	0.59
	CTREE	52	11	0.80	0.61	0.79	0.61
	Logistic reg.	34	20	0.80	0.63	0.80	0.63

Abbreviations: AUC, area under the receiving operator characteristic curve; CART, classification and regression tree; CESD-10, Center for Epidemiologic Studies Depression 10-item scale; CTREE, conditional inference tree; GAD-7, Generalized Anxiety Disorder 7-item scale; FS, flourishing scale; pCA, percent classification accuracy; reg., regression; R^2_{adj} , adjusted R^2 ; RMSE, root mean square error.

adolescent anxiety¹¹ and school connectedness to emotional distress and depression in youth.^{8,9} Additionally, previous classification tree analysis on adolescent girls found poor school functioning to be a major risk factor for depression onset, but found that parental support was only protective among subgroups with low depression at baseline.¹⁸ The protective association to school connectedness highlights the role of the school environment in helping to shape youth mental health and highlights why schools are an appropriate context for intervening, given the ability to reach a large section of the youth population. The decision tree method highlighted in the current study is well suited to future research evaluating complex environmental characteristics and co-occurring interventions.

As previously mentioned, an advantage of decision trees is the ability to examine complex interactions between predictors and identify high-risk subgroups to whom prevention and intervention efforts can be targeted. In the illustrative example with anxiety, bullying was significantly associated with the odds of having clinically relevant anxiety symptoms in the regression model; however, in the CART model, bullying only appears as a risk factor for

higher anxiety among the subset of female students with a happy home life and lower school connectedness.

Similarly, sleep time was associated with greater odds of anxiety in the regression model, though the magnitude was small; in contrast, the CART model found sleep to be a protective factor among females without a happy home life and with high school connectedness. Estimates in the regression model correspond to the overall average association across the entire sample and do not provide any insight into the differential impacts on various subgroups. In this case, the low effect size for sleep time in the regression model masks its importance among a specific subgroup.

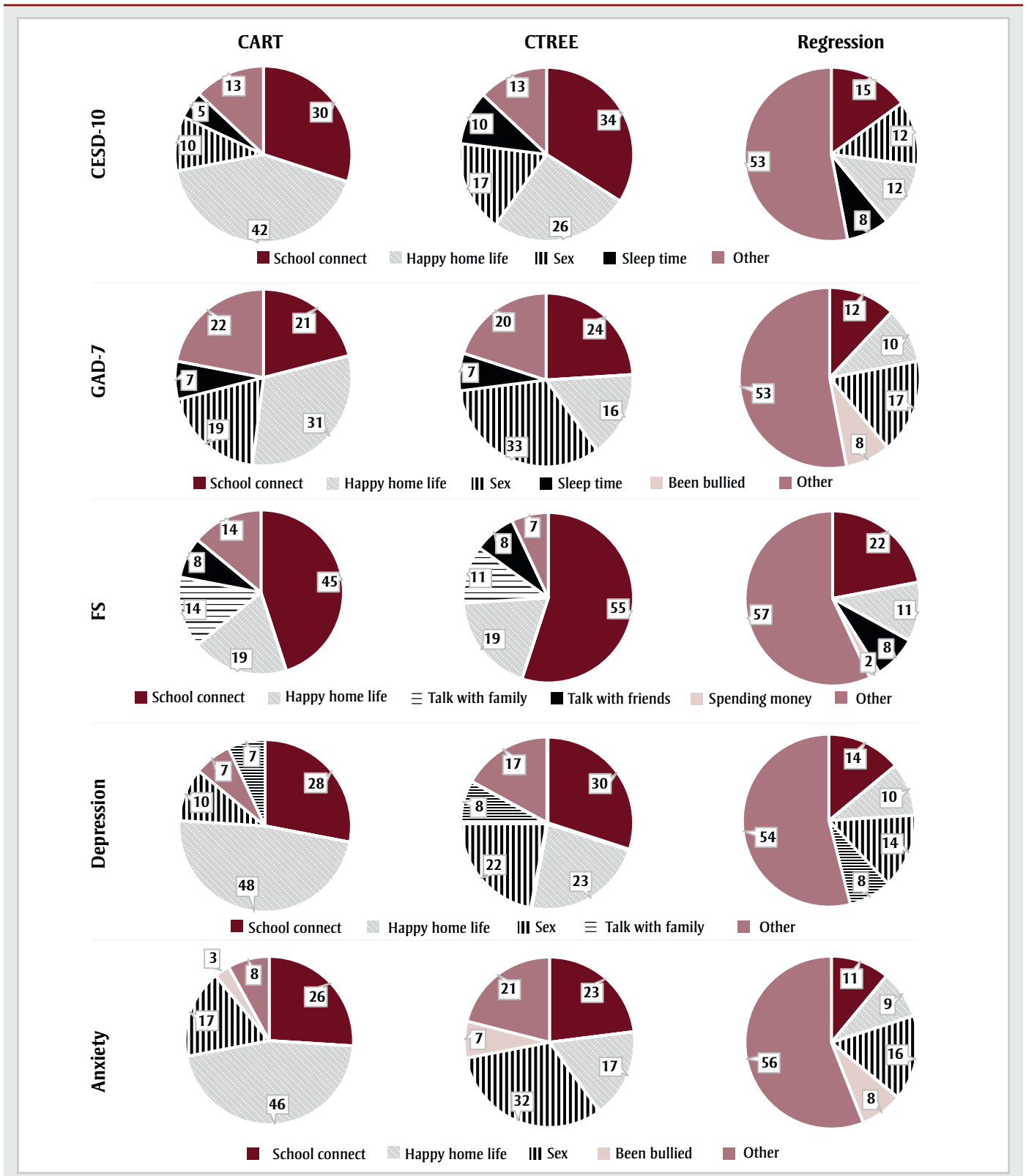
Studies by Handley et al.²⁷ and Batterham et al.,²⁸ which examined suicide ideation in adults, each found important factors present in decision tree analyses that were not significant in corresponding regression models. As noted by Handley, this suggests a multiplicative rather than independent impact of these factors, which would not be detected using a standard regression model of main effects. Thus, decision trees can be much more useful than regression models for researchers and practitioners seeking to identify

unique characteristics of the highest risk groups to whom to tailor interventions.

Despite these findings, the stronger predictive performance of regression models compared to decision tree models seen in this study could suggest that the underlying nature of predictors is somewhat linear. In the illustrative anxiety example, school connectedness was found to be an important factor across both happy home life subgroups, while sex was found to be the next most important factor across three of four subsequent subgroups. This suggests that the effect of these factors is similar across the sample, meaning a regression analysis would adequately capture this effect through the single model estimate. Decision trees have a greater advantage over regression models when the true underlying relationships in the data are nonlinear.¹² Researchers should therefore carefully consider underlying data structures based on theory and descriptive exploration when contemplating the most appropriate analysis technique.

This study examined two types of decision trees: CART and CTREE. Both models segment the population into distinct subgroups by recursively choosing the variables and cut-off levels that produce

FIGURE 2
Relative variable importance percentages of top contributing predictor variables for CART, CTREE and regression models for continuous and binary outcomes



Abbreviations: CART, classification and regression tree; CESD-10, Center for Epidemiologic Studies Depression 10-item scale; CTREE, conditional inference tree; GAD-7, Generalized Anxiety Disorder 7-item scale; FS, flourishing scale; school connect, school connectedness.

maximum separation among subgroups and minimal within-group variability. While CART and CTREE performed similarly in terms of prediction accuracy, CART consistently produced more parsimonious models, including fewer total model parameters and unique variables. Both CART and CTREE models tended to include multiple splits on different values of the same variable, particularly for the continuous outcomes examined. Tendency to favour continuous predictors over categorical due to the greater number of potential splits is a commonly noted drawback of decision trees.^{12,31} For binary outcomes, this limitation seems to be more of a concern for CTREE than CART.

Another commonly mentioned drawback of decision trees is the tendency for the models to overfit to the sample data,³⁵ which is partially mitigated by pruning in the case of CART and stopping rules based on tests of statistical significance in the case of CTREE.³⁵ In this study, similar model performance for training and test sets showed that overfitting is not a concern using either method, which may potentially be credited to the large sample size in this dataset. Interestingly, CTREE produced much more complex models than CART. CTREE models in this study used a standard statistical significance threshold of $\alpha = 0.05$ with Bonferroni correction, suggesting that perhaps more stringent criteria should be used with CTREE in the case of large sample size. Thus, while previous literature tends to favour CTREE,¹² this study suggests that researchers working with larger-scale health data should instead consider using CART when parsimony and interpretability are primary concerns.

Strengths and limitations

This study provides a novel application of decision trees using large-scale Canadian health survey data. In contrast to previous limited research, this study benefits from a large sample size that allows for more complex tree structures.

However, the resulting increased tree complexity makes interpretation difficult, which diminishes one of the primary benefits of tree analysis. While this study used standard stopping and pruning criteria, additional restrictions such as limiting the number of levels and using more stringent significance thresholds could produce smaller, more easily interpretable trees.

The impact of varying restrictions on overall model fit should be tested in future work. Additionally, only main effects were included in the regression models for this study; inclusion of interaction terms could have increased the relative performance, though as previously noted this can lead to issues in computation and interpretation.

Another limitation of this study is the low overall model fit. Test set R^2_{adj} values for continuous outcomes ranged from 0.27 to 0.51, indicating that the included predictors explain less than half of the overall variation in the outcomes. AUCs for binary outcomes ranged from 0.59 to 0.70, indicating low to moderate discriminative ability. While it is not uncommon for behavioural studies to have lower model fits, this suggests that other intrinsic factors that are not captured in this study may play an important role in predicting mental health outcomes. Previous studies of suicide ideation outcomes have generally seen higher AUCs around 0.80;^{15,21,27} however, these studies included baseline depression, which is already a well-established predictor.

Additionally, this study uses a cross-sectional, nonrandomized study design, meaning that neither decision trees nor regression models can show causal relationships between the predictors and mental health outcomes in this case. More broadly, decision trees are generally considered to be exploratory methods¹² used for hypothesis generation. Further, decision trees are not deterministic methods and are highly sensitive to the sample and parameter choices. Methods such as random forest, which grow multiple trees and aggregate the results into overall measures of variable importance, have been developed to overcome this instability,⁴⁶ though interpretability is sacrificed. Finally, the CART and CTREE methods used in this study do not account for the hierarchical nature of data (i.e. students clustered within schools). Newer tree methods such as RE-EM^{47,48} and M-CART⁴⁹ have been developed to account for this nonindependence of observations and should be examined in future research.

Conclusion

Despite growing use in other domains, decision trees remain an underutilized analysis technique in population health research. While the predictive performance of decision trees was found to be

slightly lower than that of traditional regression methods, trees provide a means of examining complex interactions between predictors, and present results in a form that is easily interpretable by nontechnical audiences, aiding in knowledge translation. The ability of decision trees to identify high-risk subgroups to whom prevention and intervention efforts can be targeted is particularly valuable to public health practitioners facing limited resources. Decision trees can be a powerful addition to population health researchers' methodology repertoire to address research questions that cannot be answered by traditional regression methods.

Acknowledgements

The COMPASS study has been supported by a bridge grant from the Canadian Institutes of Health Research (CIHR) Institute of Nutrition, Metabolism and Diabetes (INMD) through the "Obesity – Interventions to Prevent or Treat" priority funding awards (OOP-110788; awarded to SL); an operating grant from the CIHR Institute of Population and Public Health (MOP-114875; awarded to SL); a CIHR project grant (PJT-148562; awarded to SL); a CIHR bridge grant (PJT-149092; awarded to KP/SL); a CIHR project grant (PJT-159693; awarded to KP); and by a research funding arrangement with Health Canada (#1617-HQ-000012; contract awarded to SL) and a CIHR-Canadian Centre on Substance Abuse team grant (OF7 B1-PCPEGT 410-10-9633; awarded to SL). The COMPASS-Quebec project additionally benefits from funding from the Ministère de la Santé et des Services sociaux of the province of Quebec, and the Direction régionale de santé publique du CIUSSS de la Capitale-Nationale. LD received funding from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2016-04396).

Conflicts of interest

Scott Leatherdale is an Associate Scientific Editor with the HPCDP Journal but has recused himself from the review process for this paper. The authors declare there are no other conflicts of interest.

Authors' contributions and statement

KB conceived this work, conducted the statistical analysis and drafted the manuscript as part of her PhD dissertation at

the University of Waterloo. SL and JD supervised KB in conceptualizing this work and drafting the manuscript. LD supported the analyses and interpretation of the data. LD and KP provided ideas and thoughts for discussion and revised the manuscript for important intellectual content. SL is the principal investigator of the COMPASS study and led study implementation. KP conceptualized the COMPASS mental health module. All authors informed the analysis plan, provided feedback on drafts and read and approved the final manuscript.

The content and views expressed in this article are those of the authors and do not necessarily reflect those of the Government of Canada.

References

1. Reiss F. Socioeconomic inequalities and mental health problems in children and adolescents: a systematic review. *Soc Sci Med*. 2013;90:24-31. <https://doi.org/10.1016/j.socscimed.2013.04.026>
2. Quek YH, Tam WWS, Zhang MWB, Ho RCM. Exploring the association between childhood and adolescent obesity and depression: a meta-analysis. *Obes Rev*. 2017;18(7):742-54. <https://doi.org/10.1111/obr.12535>
3. Khalid S, Williams CM, Reynolds SA. Is there an association between diet and depression in children and adolescents? A systematic review. *Br J Nutr*. 2016;116(12):2097-108. <https://doi.org/10.1017/s0007114516004359>
4. Rodriguez-Ayllon M, Cadenas-Sánchez C, Estévez-López F, et al. Role of physical activity and sedentary behavior in the mental health of preschoolers, children and adolescents: a systematic review and meta-analysis. *Sports Med*. 2019;49(9):1383-410. <https://doi.org/10.1007/s40279-019-01099-5>
5. Zhang J, Paksarian D, Lamers F, Hickie IB, He J, Merikangas KR. Sleep patterns and mental health correlates in US adolescents. *J Pediatr*. 2017;182:137-43. <https://doi.org/10.1016/j.jpeds.2016.11.007>
6. Leadbeater BJ, Ames ME, Linden-Carmichael AN. Age-varying effects of cannabis use frequency and disorder on symptoms of psychosis, depression and anxiety in adolescents and adults. *Addiction*. 2019;114(2):278-93. <https://doi.org/10.1111/add.14459>
7. Romano I, Butler A, Patte KA, Ferro MA, Leatherdale ST. High school bullying and mental disorder: an examination of the association with flourishing and emotional regulation. *Int J Bullying Prev*. 2020;2(4):241-52. <https://doi.org/10.1007/s42380-019-00035-5>
8. Wilkinson-Lee AM, Zhang Q, Nuno VL, Wilhelm MS. Adolescent emotional distress: the role of family obligations and school connectedness. *J Youth Adolesc*. 2011;40(2):221-30. <https://doi.org/10.1007/s10964-009-9494-9>
9. Millings A, Buck R, Montgomery A, Spears M, Stallard P. School connectedness, peer attachment, and self-esteem as predictors of adolescent depression. *J Adolesc*. 2012;35(4):1061-7. <https://doi.org/10.1016/j.adolescence.2012.02.015>
10. Roach A. Supportive peer relationships and mental health in adolescence: an integrative review. *Issues Ment Health Nurs*. 2018;39(9):723-37. <https://doi.org/10.1080/01612840.2018.1496498>
11. Stuart Parrigon KL, Kerns KA. Family processes in child anxiety: the long-term impact of fathers and mothers. *J Abnorm Child Psychol*. 2016;44(7):1253-66. <https://doi.org/10.1007/s10802-015-0118-4>
12. Venkatasubramaniam A, Wolfson J, Mitchell N, Barnes T, JaKa M, French S. Decision trees in epidemiological research. *Emerg Themes Epidemiol*. 2017;14(11). <https://doi.org/10.1186/s12982-017-0064-4>
13. Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W. Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. *Ann Behav Med*. 2003;26(3):172-81. https://doi.org/10.1207/s15324796abm2603_02
14. Bai Z, Xu Z, Xu X, Qin X, Hu W, Hu Z. Association between social capital and depression among older people: evidence from Anhui Province, China. *BMC Public Health*. 2020;20(1):1560. <https://doi.org/10.1186/s12889-020-09657-7>
15. Mitsui N, Asakura S, Takanobu K, et al. Prediction of major depressive episodes and suicide-related ideation over a 3-year interval among Japanese undergraduates. *PLOS ONE*. 2018;13(7):e0201047. <https://doi.org/10.1371/journal.pone.0201047>
16. Hill RM, Pettit JW, Lewinsohn PM, Seeley JR, Klein DN. Escalation to major depressive disorder among adolescents with subthreshold depressive symptoms: evidence of distinct subgroups at risk. *J Affect Disord*. 2014;158:133-8. <https://doi.org/10.1016/j.jad.2014.02.011>
17. Batterham PJ, Christensen H, Mackinnon AJ. Modifiable risk factors predicting major depressive disorder at four year follow-up: a decision tree approach. *BMC Psychiatry*. 2009;9:75. <https://doi.org/10.1186/1471-244X-9-75>
18. Seeley JR, Stice E, Rohde P. Screening for depression prevention: identifying adolescent girls at high risk for future depression. *J Abnorm Psychol*. 2009;118(1):161-70. <https://doi.org/10.1037/a0014741>
19. Smits F, Smits N, Schoevers R, Deeg D, Beekman A, Cuijpers P. An epidemiological approach to depression prevention in old age. *Am J Geriatr Psychiatry*. 2008;16(6):444-53. <https://doi.org/10.1097/jgp.0b013e3181662ab6>
20. Bae S-M. The prediction model of suicidal thoughts in Korean adults using decision tree analysis: a nationwide cross-sectional study. *PLOS ONE*. 2019;14(10):e0223220. <https://doi.org/10.1371/journal.pone.0223220>
21. Burke TA, Jacobucci R, Ammerman BA, et al. Identifying the relative importance of non-suicidal self-injury features in classifying suicidal ideation, plans, and behavior using exploratory data mining. *Psychiatry Res*. 2018;262:175-83. <https://doi.org/10.1016/j.psychres.2018.01.045>

22. Xu Y, Wang C, Shi M. Identifying Chinese adolescents with a high suicide attempt risk. *Psychiatry Res.* 2018;269:474-80. <https://doi.org/10.1016/j.psychres.2018.08.085>
23. Dykxhoorn J, Hatcher S, Roy-Gagnon M-H, Colman I. Early life predictors of adolescent suicidal thoughts and adverse outcomes in two population-based cohort studies. *PLOS ONE.* 2017;12(8):e0183182. <https://doi.org/10.1371/journal.pone.0183182>
24. Hill RM, Oosterhoff B, Kaplow JB. Prospective identification of adolescent suicide ideation using classification tree analysis: models for community-based screening. *J Consult Clin Psychol.* 2017;85(7):702-11. <https://doi.org/10.1037/ccp0000218>
25. Kim HK, Kim JY, Kim JH, Hyoung HK. Decision tree identified risk groups with high suicidal ideation in South Korea: a population-based study. *Public Health Nurs.* 2016;33(2):99-106. <https://doi.org/10.1111/phn.12219>
26. Bae SM, Lee SA, Lee S-H. Prediction by data mining, of suicide attempts in Korean adolescents: a national study. *Neuropsychiatr Dis Treat.* 2015;11:2367-75. <https://doi.org/10.2147/NDT.S91111>
27. Handley TE, Hiles SA, Inder KJ, et al. Predictors of suicidal ideation in older people: a decision tree analysis. *Am J Geriatr Psychiatry.* 2014;22(11):1325-35. <https://doi.org/10.1016/j.jagp.2013.05.009>
28. Batterham PJ, Christensen H. Longitudinal risk profiling for suicidal thoughts and behaviours in a community cohort using decision trees. *J Affect Disord.* 2012;142(1-3):306-14. <https://doi.org/10.1016/j.jad.2012.05.021>
29. Leatherdale ST, Brown KS, Carson V, et al. The COMPASS study: a longitudinal hierarchical research platform for evaluating natural experiments related to changes in school-level programs, policies and built environment resources. *BMC Public Health.* 2014;14:331. <https://doi.org/10.1186/1471-2458-14-331>
30. Loh W-Y. Fifty years of classification and regression trees. *Int Stat Rev.* 2014;82(3):329-48. <https://doi.org/10.1111/insr.12016>
31. Breiman L, Friedman J, Olshen R, et al. *Classification and regression trees.* 1st ed. Pacific Grove (CA): Wadsworth International Group; 1984. 368 p.
32. Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating.* New York (NY): Springer; 2009. 525 p.
33. James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning.* New York (NY): Springer; 2013. 440 p.
34. Therneau TM, Atkinson EJ. *An introduction to recursive partitioning using the RPART routines.* Rochester (NY): Mayo Clinic Division of Biostatistics; 2018.
35. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. *J Comput Graph Stat.* 2006;15(3):651-74. <https://doi.org/10.1198/106186006X133933>
36. Van Dam NT, Earleywine M. Validation of the Center for Epidemiologic Studies Depression Scale-Revised (CESD-R): pragmatic depression assessment in the general population. *Psychiatry Res.* 2011;186(1):128-32. <https://doi.org/10.1016/j.psychres.2010.08.018>
37. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas.* 1977;1(3):385-401. <https://doi.org/10.1177/014662167700100306>
38. Bradley K, Bagnell A, Brannen C. Factorial validity of the Center for Epidemiological Studies Depression 10 in adolescents. *Issues Ment Health Nurs.* 2010;31(6):408-12. <https://doi.org/10.3109/01612840903484105>
39. Spitzer RL, Kroenke K, Williams JW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med.* 2006;166(10):1092-7. <https://doi.org/10.1001/archinte.166.10.1092>
40. Mossman S, Luft M, Schroeder H, et al. The Generalized Anxiety Disorder 7-item scale in adolescents with generalized anxiety disorder: signal detection and validation. *Ann Clin Psychiatry.* 2017;29(4):227-34A.
41. Hone L, Jarden A, Schofield G. Psychometric properties of the flourishing scale in a New Zealand sample. *Soc Indic Res.* 2014;119(2):1031-45. <https://doi.org/10.1007/s11205-013-0501-x>
42. Howell A, Buro K. Measuring and predicting student well-being: further evidence in support of the flourishing scale and the scale of positive and negative experiences. *Soc Indic Res.* 2015;121(3):903-15. <https://doi.org/10.1007/s11205-014-0663-1>
43. Statistics Canada. GeoSearch [Internet]. Ottawa (ON): Statistics Canada; 2017 [cited 2019 Oct 2]. Available from: <https://www12.statcan.gc.ca/census-recensement/2016/geo/geosearch-georecherche/index-eng.cfm>
44. Statistics Canada. Download, Census Profile, 2016 Census [Internet]. Ottawa (ON): Statistics Canada; 2017 [cited 2019 Oct 2]. Available from: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/download-telecharger/comp/page_dl-tc.cfm?Lang=E
45. Mela CF, Kopalle PK. The impact of collinearity on regression analysis: the asymmetric effect of negative and positive correlations. *Appl Econ.* 2002;34(6):667-77. <https://doi.org/10.1080/00036840110058482>
46. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychol Methods.* 2009;14(4):323-48. <https://doi.org/10.1037/a0016973>
47. Finch WH. Recursive partitioning in the presence of multilevel data. *Gen Linear Model J.* 2015;41(2):30-44.
48. Sela RJ, Simonoff JS. RE-EM trees: a data mining approach for longitudinal and clustered data. *Mach Learn.* 2012;86(2):169-207. <https://doi.org/10.1007/s10994-011-5258-3>

-
49. Lin S, Luo W. A new multilevel CART algorithm for multilevel data with binary outcomes. *Multivariate Behav Res.* 2019;54(4):578-92. <https://doi.org/10.1080/00273171.2018.1552555>