

Recherche quantitative originale

Utilisation des arbres décisionnels dans la recherche en surveillance de la santé de la population : application aux données d'enquête sur la santé mentale des jeunes de l'étude COMPASS

Katelyn Battista, M. Math (1); Liqun Diao, Ph. D. (2); Karen A. Patte, Ph. D. (3); Joel A. Dubin, Ph. D. (1,2); Scott T. Leatherdale, Ph. D. (1)

Cet article a fait l'objet d'une évaluation par les pairs.

 Diffuser cet article sur Twitter

Résumé

Introduction. Dans la recherche en surveillance de la santé de la population, les données d'enquête sont couramment analysées à l'aide de méthodes de régression. Or ces méthodes disposent d'une capacité limitée à analyser les relations complexes. À l'opposé, les modèles d'arbres décisionnels sont parfaitement adaptés pour segmenter les populations et étudier les interactions complexes entre facteurs, et leur utilisation dans la recherche en santé est en pleine croissance. Cet article fournit un aperçu de la méthodologie des arbres décisionnels et de leur application aux données d'enquête sur la santé mentale des jeunes.

Méthodologie. La performance de deux techniques courantes de construction d'arbres décisionnels, soit l'arbre de classification et de régression (CART) et l'arbre d'inférence conditionnelle (CTREE), est comparée aux modèles classiques de régression linéaire et logistique par le biais d'une application aux résultats en santé mentale des jeunes de l'étude COMPASS. Les données ont été recueillies auprès de 74 501 élèves de 136 écoles au Canada. L'anxiété, la dépression et le bien-être psychologique ont été mesurés, de même que 23 variables sociodémographiques et facteurs de prédiction des comportements liés à la santé. La performance du modèle a été évaluée au moyen de mesures de prédiction de l'exactitude, de la parcimonie et de l'importance relative des variables.

Résultats. Les modèles d'arbres décisionnels et les modèles de régression ont systématiquement mis en évidence les mêmes ensembles de facteurs de prédiction les plus importants pour chaque résultat, ce qui indique un niveau général de concordance entre méthodes. Trois modèles ont présenté une exactitude prédictive plus faible, mais se caractérisent par une plus grande parcimonie et accordent une importance relative plus élevée aux principaux facteurs de différenciation.

Conclusion. Les arbres décisionnels permettent de cerner les sous-groupes à risque élevé qu'il convient de cibler dans le cadre des efforts de prévention et d'intervention. Ils constituent donc un outil utile pour répondre aux questions de recherche auxquelles les méthodes de régression classiques ne peuvent pas répondre.

Mots-clés : *arbres décisionnels, santé de la population, méthodes d'enquête, santé mentale, jeunes*

Points saillants

- Les arbres décisionnels sont utilisables en recherche en santé de la population pour répondre à des questions de recherche auxquelles les méthodes de régression classiques ne peuvent pas répondre.
- Un des principaux avantages des arbres décisionnels par rapport aux modèles de régression est qu'ils permettent d'étudier les interactions complexes entre facteurs de risque.
- Les arbres décisionnels sont utilisables pour cerner les groupes à risque élevé qu'il convient de cibler dans le cadre des efforts de prévention et d'intervention.
- Bien que les modèles de régression puissent présenter une exactitude prédictive supérieure dans certains contextes, les arbres décisionnels accordent une plus grande importance aux principaux facteurs de différenciation.

Introduction

La recherche en surveillance de la santé de la population est souvent réalisée à l'aide d'enquêtes à grande échelle qui tentent d'évaluer les répercussions d'un large éventail de facteurs sociaux, économiques et environnementaux sur divers résultats de santé. La relation entre ces

Rattachement des auteurs :

1. École des sciences de la santé publique, Université de Waterloo, Waterloo (Ontario), Canada
2. Département de statistique et d'actuariat, Université de Waterloo, Waterloo (Ontario), Canada
3. Département des sciences de la santé, Université Brock, St. Catharines (Ontario), Canada

Correspondance : Katelyn Battista, École des sciences de la santé publique, Université de Waterloo, 200, avenue University Ouest, Waterloo (Ontario) N2L 3G1; tél. : 519-888-4567, poste 46706; courriel : kbattista@uwaterloo.ca

facteurs et les résultats de santé est souvent caractérisée par des interactions complexes qui font en sorte qu'il est impossible d'isoler un facteur de causalité unique. Dans le contexte de la santé mentale des jeunes, des études antérieures ont établi que les résultats sont associés au statut socioéconomique¹, au poids², aux comportements alimentaires³, à l'activité physique et aux comportements sédentaires⁴, aux habitudes de sommeil⁵, à la consommation de cannabis⁶, à l'intimidation⁷, au sentiment d'appartenance à l'école^{8,9} et aux relations avec les pairs et la famille^{10,11}, entre autres facteurs. Cependant, la plupart des études de recherche portent sur les répercussions de tout facteur ou domaine de facteurs donné pris isolément, alors qu'en fait, les interrelations sous-jacentes sont vraisemblablement plus complexes.

Les liens sont souvent étudiés à l'aide de modèles de régression, qui fournissent une estimation de l'association entre un prédicteur et un résultat tout en tenant compte d'autres facteurs. Cependant, ces modèles sont rarement utilisés pour estimer les interactions complexes entre facteurs, en raison des contraintes liées aux calculs et des difficultés d'interprétation. De plus, les estimations du modèle résultant ne permettent pas d'élaborer les profils de risque, c'est-à-dire de séparer les sujets en sous-groupes en fonction de certaines combinaisons de facteurs de risque. Or il importe de cerner les sous-groupes à risque élevé pour cibler efficacement les ressources et les interventions. Les arbres décisionnels forment une classe de modèles différente qui est parfaitement adaptée pour segmenter les populations et étudier les interactions complexes entre facteurs¹².

Les arbres décisionnels sont couramment utilisés en recherche clinique axée sur le dépistage et le diagnostic¹³, en particulier sur la prévision. Leur utilisation, même si elle connaît une hausse, est moins fréquente dans la recherche en santé de la population, qui est axée sur la compréhension des associations et l'identification des sous-groupes à cibler dans le cadre des interventions comportementales. Dans le domaine de la santé mentale, les études récentes utilisant des arbres décisionnels ont principalement analysé les liens avec la dépression¹⁴⁻¹⁹ et le risque de suicide^{15,20-28}.

Deux études ont analysé les résultats liés à la dépression chez des populations de

jeunes en particulier. Hill et ses collaborateurs¹⁶ ont découvert que, chez les élèves présentant des symptômes dépressifs sous-cliniques au début de l'étude, le soutien d'amis avait un effet protecteur contre la survenue d'un trouble dépressif majeur avant l'âge de 30 ans, et que le risque de trouble d'anxiété et de trouble lié à la consommation de substances augmentait chez les personnes qui ne pouvaient pas compter sur le soutien d'amis. Seeley, Stice et Rohde¹⁸ ont constaté qu'un fonctionnement scolaire déficient constituait l'un des principaux facteurs de risque associés à l'apparition d'un trouble dépressif majeur chez les filles présentant des symptômes dépressifs importants au début de l'étude, le soutien parental agissant comme un facteur de protection uniquement chez les filles présentant des symptômes dépressifs légers au début de l'étude. Trois études ont porté sur les idées suicidaires chez les jeunes et ont constaté que les facteurs intermédiaires comme les relations familiales^{22,26} et le soutien social^{22,24} constituaient des facteurs de protection uniquement chez les élèves ne présentant pas un niveau élevé de dépression.

Parmi les études susmentionnées, quelques-unes ont intégré des comparaisons directes de la performance des arbres décisionnels par rapport aux modèles de régression. Des études de moindre envergure réalisées par Burke et ses collaborateurs²¹, Mitsui et ses collaborateurs¹⁵ et Handley et ses collaborateurs²⁷ ont révélé que les modèles de régression présentaient une exactitude prédictive supérieure aux modèles d'arbres décisionnels correspondants, mais ces études portaient sur des échantillons de petite taille (entre 359 et 2194 participants). À l'inverse, deux études à plus grande échelle – l'une menée par Dykxhoorn et ses collaborateurs²³ auprès d'un échantillon longitudinal de 11 088 enfants, et l'autre par Batterham et ses collaborateurs¹⁷ auprès d'un échantillon longitudinal de 6605 adultes – ont révélé que les arbres décisionnels surpassaient les analyses de régression logistique quant à la sensibilité et à l'exactitude prédictive globale. Si certaines données probantes indiquent que les arbres décisionnels sont susceptibles de présenter des avantages par rapport aux méthodes de régression classiques dans le cas des échantillons de grande taille, les données sont de manière générale insuffisantes dans le domaine de la santé mentale.

Malgré l'utilisation croissante des arbres décisionnels, les modèles de régression demeurent courants dans la littérature en santé de la population. Il en résulte une occasion manquée de comprendre les interactions complexes entre facteurs de risque et d'identifier les sous-groupes à risque élevé à cibler dans le cadre des efforts de prévention et d'intervention. Cette étude a donc eu pour but d'étudier l'utilisation des arbres décisionnels dans l'analyse des données en surveillance de la santé de la population à grande échelle. Dans cet article, nous fournissons d'abord un aperçu de deux types d'arbres décisionnels courants, c'est-à-dire l'arbre de classification et de régression (CART) et l'arbre d'inférence conditionnelle (CTREE). Ensuite, nous comparons la performance des modèles d'arbres décisionnels aux modèles de régression linéaire et logistique classiques par le biais d'une application aux résultats sur la santé mentale des jeunes de l'étude COMPASS²⁹. Les méthodes d'arborescence et de régression ont été évaluées en fonction de leur exactitude prédictive et de leur parcimonie, ainsi que de certains éléments concernant l'importance relative des variables et de la facilité d'interprétation du modèle.

Méthodologie

Renseignements généraux sur les arbres décisionnels

Les arbres décisionnels sont des modèles statistiques qui portent sur un résultat d'intérêt en divisant l'échantillon en sous-groupes en fonction de variables prédictives. Les sous-groupes sont déterminés à l'aide d'une série de divisions binaires qui ressemblent à une structure arborescente. Divers types d'algorithmes d'arbres décisionnels ont été élaborés³⁰. Notre analyse porte sur deux types d'arbres décisionnels courants : CART et CTREE. Un aperçu de la méthodologie des arbres CART et CTREE dans le contexte de la recherche épidémiologique a déjà été publié^{12,13}. Nous présentons ici un résumé de leurs principales caractéristiques.

Arbres de classification et de régression

Les arbres CART sont un type d'arbre décisionnel qui est largement utilisé à la fois pour les résultats catégoriels (classification) et pour les résultats continus (régression). Élaborées initialement par Breiman et ses collaborateurs³¹, les méthodes CART permettent de sélectionner les divisions optimales dans l'échantillon afin

d'obtenir des sous-groupes³² formés de sujets similaires et de faire en sorte que les sujets appartenant à des sous-groupes distincts soient aussi différents que possible. Les divisions optimales sont déterminées grâce à une sélection récursive des variables et à des seuils générant une séparation maximale entre les sous-groupes et une variabilité minimale au sein d'un même groupe en matière de résultats³². Les variables continues et catégorielles peuvent être divisées à plusieurs reprises dans l'arbre à différents seuils. La segmentation se poursuit jusqu'à ce qu'un critère d'arrêt soit atteint, lequel est généralement fondé sur la taille minimale du sous-groupe^{12,32,33}. Au moyen de ce processus récursif, l'espace prédictif est segmenté en un ensemble de sous-groupe finaux, pour chacun desquels la valeur moyenne du résultat (arbres de régression) ou le pourcentage du sous-groupe associé au résultat (arbres de classification) est calculé³³.

Un arbre de grande taille construit au moyen de la division récursive de l'espace prédictif a tendance à se surajuster aux données de l'échantillon, de sorte que les résultats sont peu généralisables. Le surajustement est atténué grâce à une procédure d'élagage et de validation croisée, où l'élagage de l'arbre de grande taille mène à une séquence de sous-arbres imbriqués à partir de laquelle un arbre optimal est retenu. La méthode la plus courante d'élagage est l'élagage par coût-complexité, selon lequel une séquence croissante de paramètres de complexité correspond à une séquence de sous-arbres imbriqués de taille décroissante^{33,34}. Le sous-arbre optimal qui réduit au minimum l'erreur moyenne en fonction de la validation croisée³³ est ensuite choisi. Lorsque les échantillons sont de plus grande taille, la règle « 1-SE » est souvent utilisée pour choisir le sous-arbre le plus petit comportant une erreur moyenne se situant à moins d'un écart-type de l'erreur minimale globale^{12,13}.

Arbres d'inférence conditionnelle

Les arbres CTREE offrent une méthodologie de rechange aux arbres CART, qui a été élaborée par Hothorn et ses collaborateurs³⁵. Alors que l'arbre CART choisit la division optimale à chaque étape parmi toutes les variables possibles et points de segmentation simultanément, l'arbre CTREE sépare la détermination de la segmentation en deux étapes. D'abord, la variable optimale à segmenter est choisie en fonction de l'association la plus forte avec les

résultats. L'association à la variable de résultat est mesurée à l'aide de modèles de régression adaptés aux résultats, par exemple, une régression linéaire dans le cas de résultats continus et une régression logistique dans le cas de résultats binaires^{12,35}. La covariable présentant la plus faible valeur p est retenue pour la segmentation. Ensuite, le point de segmentation optimal pour cette variable est déterminé^{12,35}. Cette approche permet d'atténuer le biais de sélection des variables comportant de nombreux points de segmentation, un biais souvent observé dans les arbres CART^{12,35}. Le processus de segmentation se poursuit de façon récursive dans chaque sous-groupe jusqu'à ce qu'un critère d'arrêt soit atteint. Comme pour l'arbre CART, des variables continues et catégorielles peuvent être segmentées plus d'une fois dans l'arbre à différents seuils.

Le critère d'arrêt pour l'arbre CTREE est fondé sur une hypothèse nulle globale : l'algorithme met fin à la segmentation lorsqu'aucune covariable ne présente d'association significative avec le résultat en fonction d'un seuil de signification pré-établi (α , α)^{12,35}. Dans le cas d'échantillons de grande taille, des critères d'arrêt supplémentaires fondés sur la taille minimale des sous-groupes sont également utilisables. Aucun élagage n'est nécessaire dans le cas de l'arbre CTREE : le test global de signification sert à prévenir le surajustement^{12,35}.

Application

Nous avons comparé la performance relative des arbres décisionnels et des méthodes de régression dans le contexte d'une recherche en surveillance de la population par le biais de données sur la santé mentale des jeunes tirées de l'étude COMPASS²⁹.

Approbation éthique, plan d'étude et échantillon

L'étude COMPASS est une étude de cohorte prospective conçue pour recueillir des données hiérarchiques sur la santé auprès des élèves des écoles secondaires du Canada²⁹. Le Comité d'éthique de la recherche de l'Université de Waterloo (BER 30118) a approuvé l'étude COMPASS. D'autres détails sur l'étude de référence COMPASS sont accessibles en version imprimée²⁹ et en ligne (<https://uwaterloo.ca/compass-system>, en anglais seulement).

Nous avons utilisé les données recueillies auprès des élèves au cours de l'année 7 (2018-2019) de l'étude COMPASS. L'échantillon se composait de 74 501 élèves de 136 écoles de l'Ontario (61 écoles), de l'Alberta (8 écoles), de la Colombie-Britannique (15 écoles) et du Québec (52 écoles). L'étude COMPASS utilise un échantillonnage dirigé pour former des échantillons représentatifs de toute l'école après utilisation de protocoles d'information active et de consentement parental passif. Le taux de participation pour 2018-2019 a été de 81,9 % et la non-participation est principalement attribuable à l'absentéisme ou aux périodes libres prévues à la date de collecte des données.

Mesures

Le questionnaire COMPASS destiné aux élèves est un questionnaire imprimé rempli par les élèves pendant les heures de classe. Le questionnaire est anonyme et auto-administré, et les élèves ont en tout temps le choix de ne pas participer. Cette étude a porté sur cinq mesures de résultats en santé mentale liées à la dépression, à l'anxiété et au bien-être psychosocial (épanouissement), ainsi que 23 mesures prédictives de base liées aux caractéristiques sociodémographiques, au poids, à une saine alimentation, aux comportements en matière de mouvement, à l'utilisation de substances, à l'intimidation, au soutien aux études et scolaire ainsi qu'à la perception du soutien offert par la famille et les amis.

Résultats en santé mentale

Dépression

La dépression a été mesurée à l'aide de l'échelle d'évaluation de la dépression du Center for Epidemiologic Studies, CESD-10) à 10 items^{36,37}, qui a été validée auprès de populations d'adolescents³⁸. L'échelle CESD-10 fournit un score sur une échelle continue allant de 0 à 30, les scores les plus élevés correspondant à des symptômes dépressifs plus importants et à un risque de dépression unipolaire. L'étude comporte également une mesure binaire de la dépression, selon laquelle les élèves obtenant un score supérieur ou égal à 10 sont considérés comme présentant des symptômes de dépression cliniquement significatifs.

Anxiété

L'anxiété a été mesurée au moyen de l'échelle d'évaluation du trouble d'anxiété généralisée (Generalized Anxiety Disorder, GAD-7) à 7 items³⁹, qui a été validée auprès

de populations d'adolescents⁴⁰. L'échelle GAD-7 fournit un score sur une échelle continue allant de 0 à 21, les scores les plus élevés correspondant à des niveaux supérieurs d'anxiété. L'étude produit également une mesure binaire de l'anxiété, selon laquelle les élèves obtenant un score supérieur ou égal à 10 sont considérés comme présentant des symptômes d'anxiété cliniquement significatifs.

Épanouissement

L'épanouissement est une composante du bien-être psychologique et il est mesuré au moyen d'une version modifiée de l'échelle d'évaluation de l'épanouissement (Diener's Flourishing Scale, ou FS)⁴¹, qui a été validée auprès des jeunes adultes⁴². L'échelle FS fournit un score sur une échelle continue allant de 8 à 40, les scores les plus élevés correspondant à des niveaux supérieurs d'épanouissement.

Variables prédictives

Les variables prédictives sociodémographiques sont l'âge, le sexe, l'origine ethnique et l'argent de poche hebdomadaire (un indicateur du statut socioéconomique). Le poids a été mesuré à l'aide de la perception du poids et de la classification de l'indice de masse corporelle (IMC). La saine alimentation a été mesurée au moyen d'un indicateur binaire précisant si les élèves prenaient un déjeuner chaque jour, et du nombre de portions de fruits et de légumes consommés quotidiennement. Les comportements en matière de mouvement ont été évalués à l'aide du nombre moyen de minutes par jour consacrées à l'activité physique d'intensité modérée à vigoureuse, du temps total en minutes passé devant un écran chaque jour et du temps consacré au sommeil par jour en minutes. La consommation de substances a été mesurée au moyen d'indicateurs binaires de tabagisme, d'utilisation de la cigarette électronique et de consommation de cannabis au cours du dernier mois ainsi que de la consommation excessive d'alcool au cours du dernier mois. L'intimidation a été mesurée au moyen de deux indicateurs, l'un mesurant si l'élève a été intimidé et l'autre mesurant si l'élève a intimidé d'autres élèves au cours des 30 derniers jours. Le soutien aux études et le soutien scolaire ont été mesurés par un indicateur binaire précisant si les élèves envisageaient de faire des études postsecondaires ou non, par le nombre de cours manqués au cours des quatre dernières semaines et par un score indiquant le sentiment d'appartenance à l'école (les scores

les plus élevés indiquant un sentiment d'appartenance supérieur à l'égard de l'école). La perception du soutien offert par la famille et les amis a été mesurée au moyen des indicateurs binaires suivants : avoir une vie familiale heureuse ou non, se sentir capable de parler de ses problèmes avec sa famille ou non et se sentir capable de parler de ses problèmes avec des ami(e)s ou non.

Outre les mesures recueillies au niveau de l'élève, d'autres prédicteurs relatifs au milieu scolaire ont été intégrés : le taux d'inscription à l'école, la province, le revenu médian au sein de la zone scolaire et l'indice d'urbanisation de l'école. Les mesures du revenu et de l'indice d'urbanisation ont été tirées du recensement de 2016 de Statistique Canada et des valeurs liées à la région de tri d'acheminement associée à l'école^{43,44}.

Outils de mesure

En ce qui concerne les éléments individuels des échelles de cotation liées à la santé mentale, une moyenne-sujet a été attribuée aux élèves pour lesquels un ou deux éléments étaient manquants. Bien que l'imputation par la moyenne puisse réduire artificiellement la variance, nous n'avons pas utilisé de méthodes d'imputation plus complexes dans la mesure où l'objectif premier était l'analyse de la performance plutôt que l'inférence. Les élèves pour lesquels des valeurs de variables étaient manquantes ou aberrantes ont été retirés, de sorte que l'échantillon final de cas complets comportait 52350 élèves. Les caractéristiques de l'échantillon sont présentées dans le tableau 1. L'échantillon a été segmenté de façon aléatoire en deux groupes, soit un échantillon d'entraînement (41 795; 80 %) et un échantillon d'essai (10 555; 20 %).

Les modèles CART et CTREE ont été réalisés pour les résultats continus (CESD-10, GAD-7, FS) et pour les résultats binaires (dépression, anxiété). L'élagage CART a été réalisé à l'aide de la validation croisée à 10 blocs et de la règle 1-SE (une fois l'erreur-type). Pour l'arbre CTREE, le seuil de signification a été établi à $\alpha = 0,05$ avec correction de Bonferroni pour les essais multiples. Compte tenu de la taille importante de l'échantillon, un critère d'arrêt supplémentaire a été inclus tant pour l'arbre CART que pour l'arbre CTREE, afin de limiter le nombre minimum d'observations par catégorie à 1 % de l'échantillon. Les modèles de régression

linéaire et logistique ont également été réalisés pour les résultats continus et les résultats binaires, incluant tous les effets principaux. Une sélection descendante des variables a été réalisée à l'aide du critère d'information d'Akaike (CIA) pour reproduire les résultats de l'élagage.

Des modèles ajustés obtenus à partir de l'échantillon d'entraînement ont été appliqués à l'échantillon d'essai. Leur performance prédictive a été comparée à l'aide de la valeur ajustée de R^2 (R^2_{aj}) et de l'écart moyen quadratique (EMQ) dans le cas des résultats continus et à l'aide du taux d'exactitude de la classification (TEC) et de l'aire sous la courbe caractéristique de la performance d'un test (ASC) dans le cas des résultats binaires. La valeur R^2_{aj} correspond au degré de variation expliqué par le modèle, ajusté en fonction du nombre de covariables, ce qui fait que la valeur R^2_{aj} diminue si l'inclusion d'une covariable donnée n'entraîne pas d'augmentation importante de la variation expliquée. L'EMQ correspond à la moyenne de l'écart quadratique entre les valeurs réelles de résultats et les valeurs prédites de résultats³³. Plus les valeurs prédites se rapprochent des valeurs réelles, plus l'EMQ est faible. Le taux d'exactitude de la classification mesure simplement le pourcentage d'observations pour lesquelles le modèle attribue correctement la valeur de résultats. L'ASC (également appelée la statistique de concordance) est une mesure de l'exactitude plus sophistiquée, qui tient compte à la fois de la sensibilité et de la spécificité du modèle³². Ces deux mesures varient entre 0 et 1, les valeurs les plus élevées correspondant à une plus grande exactitude du modèle.

La parcimonie a été évaluée à l'aide du nombre de paramètres et de variables uniques dans le modèle. Les mesures de l'importance relative des variables ont été calculées à partir de la diminution de l'ajustement du modèle découlant du retrait d'une variable donnée dans chaque modèle. Pour les arbres décisionnels, il s'agit de la somme des scores de validité de la division pour toutes les occurrences où la variable est utilisée pour la division optimale ou la division de substitution. Dans le cas des modèles de régression linéaires et logistiques, cette valeur est mesurée respectivement par l'augmentation de la valeur R^2_{aj} et par celle de l'ASC.

Toutes les analyses ont été effectuées au moyen de la version 4.0.3 du logiciel R

TABLEAU 1
Caractéristiques de l'échantillon d'élèves pour l'année 7 (2018-2019) de l'étude COMPASS

Catégorie	Variable	Valeur	Effectifs (n)	Pourcentage (%)	
Total			52 350	100,0 %	
Résultats en matière de santé mentale	Échelle CESD-10	Moyenne (ET)	8,50 (5,85)		
	Échelle GAD-7	Moyenne (ET)	6,02 (5,31)		
	Échelle d'évaluation de l'épanouissement	Moyenne (ET)	32,42 (5,39)		
	Dépression	Non		33 778	64,5 %
		Oui		18 572	35,5 %
	Anxiété	Non		40 568	77,5 %
		Oui		11 782	22,5 %
Facteurs sociodémographiques	Sexe	Féminin	27 483	52,5 %	
		Masculin	24 867	47,5 %	
	Âge (en années)	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 %	
		18	2 147	4,1 %	
	Origine ethnique	Blanche		37 370	71,4 %
		Noire		1 565	3,0 %
		Asiatique		5 559	10,6 %
		Latino-américaine		1 235	2,4 %
	Argent de poche	Autre/mixte		6 621	12,6 %
		0 \$		8 099	15,5 %
		1 \$ à \$20		12 701	24,3 %
		21 \$ à 40 \$		5 796	11,1 %
		41 \$ à 100 \$		6 469	12,4 %
	Province	Plus de 100 \$		10 067	19,2 %
		Ne sait pas		9 218	17,6 %
		Alberta		2 222	4,2 %
		Colombie-Britannique		7 298	13,9 %
		Ontario		20 450	39,1 %
Indice d'urbanisation	Québec		22 380	42,8 %	
	Grand centre urbain		28 684	54,8 %	
	Centre urbain de taille moyenne		5 044	9,6 %	
Revenu médian au sein de la zone scolaire (en milliers de dollars canadiens [\$ CA])	Petit centre urbain/milieu rural		18 622	35,6 %	
	Moyenne (ET)		67,33 (17,47)		
Taille de l'école (en centaines d'élèves)	Moyenne (ET)		8,49 (3,52)		

Suite à la page suivante

TABEAU 1 (suite)
Caractéristiques de l'échantillon d'élèves pour l'année 7 (2018-2019) de l'étude COMPASS

Catégorie	Variable	Valeur	Effectifs (n)	Pourcentage (%)
Poids et comportements alimentaires	Perception du poids	Poids insuffisant	8 300	15,9 %
		Poids plutôt idéal	31 877	60,9 %
		Surpoids/obésité	12 173	23,3 %
	Classification de l'IMC	Poids insuffisant	985	1,9 %
		Poids normal	29 932	57,2 %
		Surpoids	6 465	12,3 %
		Obésité	2 843	5,4 %
		Non indiqué	12 125	23,2 %
	Prend un déjeuner chaque jour	Non	25 373	48,5 %
		Oui	26 977	51,5 %
Comportements en matière de mouvement	Portions de fruits et de légumes	Moyenne (ET)	2,98 (1,93)	
	Activité physique quotidienne moyenne (min)	Moyenne (ET)	96,40 (62,14)	
	Temps d'écran (min)	Moyenne (ET)	350,97 (178,28)	
	Temps de sommeil (min)	Moyenne (ET)	451,94 (74,78)	
Consommation de substances	Tabagisme	Non	49 349	94,3 %
		Oui	3 001	5,7 %
	Utilisation de la cigarette électronique	Non	38 570	73,7 %
		Oui	13 780	26,3 %
	Consommation excessive d'alcool	Non	44 020	84,1 %
		Oui	8 330	15,9 %
	Consommation de cannabis	Non	46 683	89,2 %
		Oui	5 667	10,8 %
Intimidation au cours des 30 derniers jours	A été intimidé	Non	46 412	88,7 %
		Oui	5 938	11,3 %
	A intimidé d'autres élèves	Non	49 702	94,9 %
		Oui	2 648	5,1 %
Soutien aux études et soutien scolaire	Envisage de faire des études postsecondaires	Non	12 380	23,6 %
		Oui	39 970	76,4 %
	Cours manqués au cours des 4 dernières semaines	0 cours	34 894	66,7 %
		1 ou 2 cours	10 634	20,3 %
		3 à 5 cours	4 246	8,1 %
		6 cours ou plus	2 576	4,9 %
	Score lié au sentiment d'appartenance à l'école	Moyenne (ET)	18,67 (3,14)	
Soutien offert par la famille et les pairs	Vie familiale heureuse	Non	10 219	19,5 %
		Oui	42 131	80,5 %
	Parle de ses problèmes avec sa famille	Non	20 770	39,7 %
		Oui	31 580	60,3 %
	Parle de ses problèmes avec des ami(e)s	Non	12 748	24,4 %
		Oui	39 602	75,6 %

Abréviations : \$ CA, dollars canadiens; CESD-10, échelle d'évaluation de la dépression à 10 items du Center for Epidemiologic Studies; ET, écart-type; GAD-7, échelle d'évaluation du trouble d'anxiété généralisée à 7 items (Generalized Anxiety Disorder); IMC, indice de masse corporelle.

(R Foundation for Statistical Computing, Vienne, Autriche). Nous avons utilisé les fonctions « rpart » (progiciel « rpart ») et « ctree » (progiciel « partykit ») pour respectivement les modèles CART et CTREE. Nous avons utilisé les fonctions « lm » et « glm » (progiciel « MASS ») pour les modèles de régression linéaire et logistique respectivement.

Résultats

Le score moyen à l'échelle CESD-10 dans l'échantillon est de 8,50 (ET = 5,85) et 33,5 % des sujets de l'échantillon ont été considérés comme présentant des symptômes de dépression cliniquement significatifs. Le score moyen à l'échelle GAD-7 est de 6,02 (ET = 5,31) et 22,5 % des sujets de l'échantillon ont été considérés comme présentant des symptômes d'anxiété cliniquement significatifs. Le score à l'échelle d'évaluation de l'épanouissement (FS) est de 32,42 (ET = 5,39).

Comparaison de l'arbre décisionnel et du modèle de régression

À titre d'exemple représentatif, nous présentons les résultats du modèle CART et du modèle de régression logistique associés au résultat binaire lié à l'anxiété. L'arbre CART final ajusté montrant les résultats binaires liés à l'anxiété est présenté sur la figure 1. Le modèle a recensé 9 sous-groupes finaux au moyen de 5 variables uniques. La variable de segmentation principale est le fait d'avoir ou non une vie familiale heureuse. Les deux sous-groupes sont ensuite segmentés en fonction du sentiment d'appartenance à l'école, bien que des seuils différents aient été utilisés. On a également procédé à une segmentation pour certains sous-groupes en ce qui concerne le sexe, le temps de sommeil et le fait d'avoir subi de l'intimidation. Le sous-groupe final le plus important, qui forme 61 % de l'échantillon, est celui des élèves ayant indiqué avoir une vie familiale heureuse et présentant un score d'au moins 17,5 en ce qui concerne le sentiment d'appartenance à l'école. Au sein de ce groupe, la probabilité de présenter des symptômes d'anxiété pertinents sur le plan clinique était de 12,7 %, soit le taux le plus faible parmi tous les groupes. Le sous-groupe le plus à risque se compose d'adolescentes ayant indiqué qu'elles n'avaient pas une vie familiale heureuse et qu'elles avaient un faible sentiment d'appartenance à l'école (< 16,25) et, au sein de ce groupe, la probabilité de

présenter des symptômes d'anxiété pertinents sur le plan clinique s'élevait à 64,6 %.

Les résultats du modèle de régression logistique pour l'anxiété sont présentés dans le tableau 2. Le modèle final après application de la sélection descendante des variables comprend 20 variables (33 paramètres). À l'instar du modèle CART, avoir une vie familiale heureuse (rapport de cotes [RC] = 0,33; IC à 95 % : 0,31 à 0,34), être de sexe masculin (RC = 0,33; IC à 95 % : 0,31 à 0,34) et avoir un sentiment d'appartenance à l'école (RC = 0,88; IC à 95 % : 0,87 à 0,89) se sont révélés être d'importants prédicteurs. D'autres facteurs, notamment l'appartenance à un groupe ethnique minoritaire, le fait de recevoir plus d'argent de poche, de vivre au Québec, de vivre dans un petit centre urbain ou en milieu rural, d'estimer avoir un poids « plutôt idéal », de prendre un déjeuner chaque jour, de consacrer plus de temps au sommeil et de se sentir capable de parler de ses problèmes avec sa famille et ses amis, étaient associés à une probabilité plus faible de présenter des symptômes d'anxiété pertinents sur le plan clinique. En revanche, un âge avancé, le fait de consommer plus de fruits et de légumes, de passer plus de temps devant un écran, le tabagisme et l'utilisation de la cigarette électronique, le fait de subir de l'intimidation, de prévoir de faire des études postsecondaires et de manquer des cours étaient associés à une probabilité plus élevée de présenter des symptômes d'anxiété pertinents sur le plan clinique.

Exactitude prédictive et parcimonie

L'exactitude prédictive des résultats continus (CESD-10, GAD-7, FS) est présentée dans le tableau 3. Les modèles de régression linéaire présentent la valeur R^2_{aj} la plus élevée de l'ensemble d'essai et l'EMQ le plus faible pour les trois résultats. La valeur R^2_{aj} et l'EMQ sont similaires pour les modèles CART et CTREE, la valeur R^2_{aj} affichant des résultats systématiquement inférieurs de 4 % à 5 % par rapport aux résultats du modèle de régression linéaire et l'EMQ affichant des résultats supérieurs de 0,13 à 0,19 par rapport aux résultats du modèle de régression linéaire. Les arbres CART sont ceux qui possèdent le nombre le moins élevé de variables uniques, suivis des arbres CTREE, les modèles de régression linéaire comprenant quant à eux deux fois plus de variables. Cependant, le nombre de paramètres finaux (correspondant

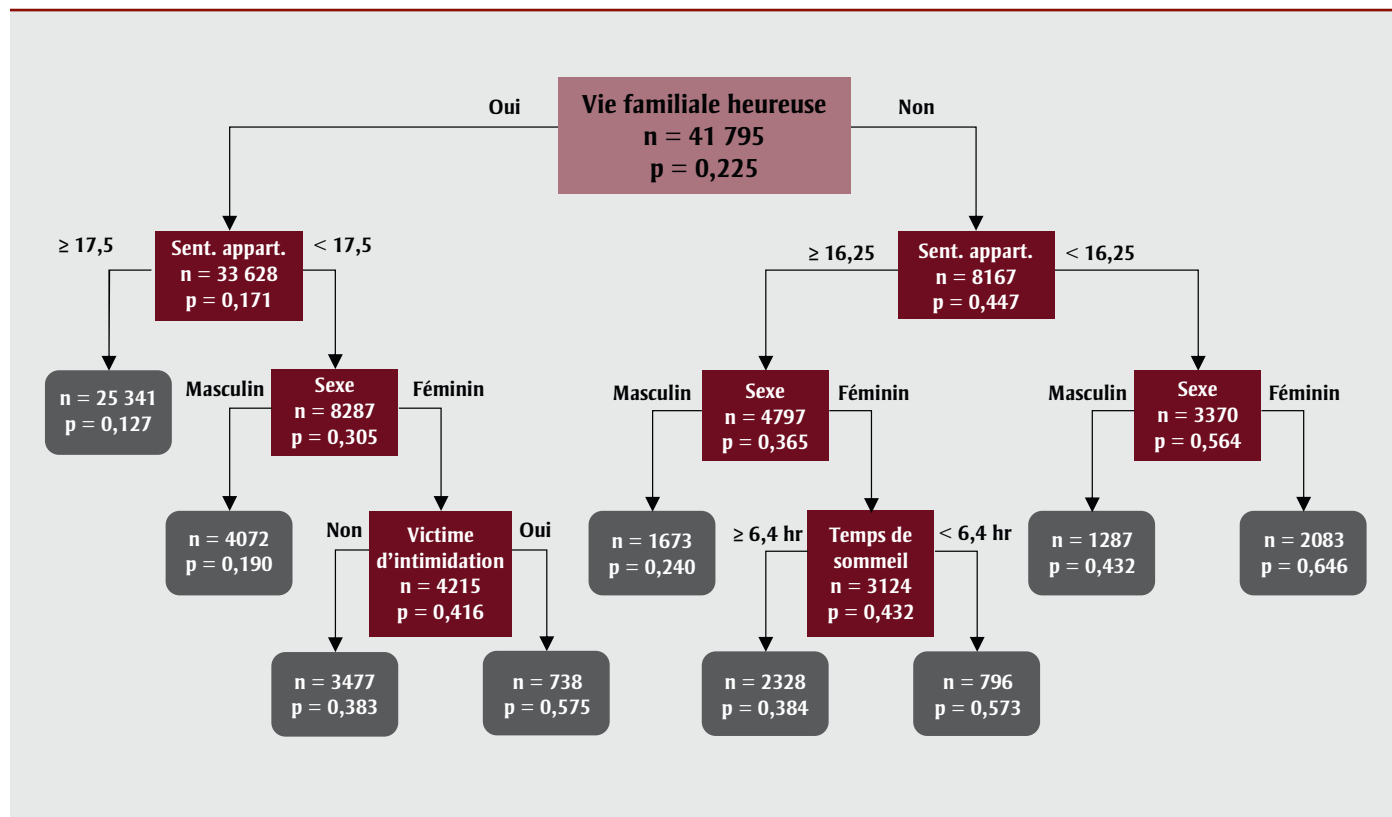
au nombre de divisions associées aux modèles d'arbre) s'est révélé semblable pour les modèles CART et les modèles de régression linéaire, et supérieur pour les modèles CTREE. La valeur absolue de R^2_{aj} est relativement faible pour tous les modèles, ce qui indique que les prédicteurs expliquent moins de la moitié de la variation du résultat. De plus, la valeur R^2_{aj} et l'EMQ calculés au moyen de l'ensemble d'essai sont similaires à ceux calculés au moyen de l'ensemble de données d'entraînement pour tous les modèles, ce qui semble indiquer un surajustement minime.

L'exactitude prédictive des résultats binaires concernant la dépression et l'anxiété est présentée dans le tableau 3. Les modèles CART sont plus parcimonieux que les modèles CTREE et les modèles de régression logistique, avec seulement 9 divisions pour les 6 variables de dépression et 8 divisions pour les 5 variables d'anxiété. Les modèles CTREE sont plus complexes et comportent plus de 50 divisions. La différence plus importante entre le nombre de sous-groupes et de variables utilisés dans les modèles CTREE comparativement aux modèles CART s'explique en partie par la segmentation du modèle en fonction des mêmes variables à plusieurs reprises à l'aide de différents seuils. Les modèles de régression logistique comprennent 22 variables pour la dépression et 20 pour l'anxiété. Malgré la différence quant à la complexité des modèles, le taux d'exactitude de la classification et l'ASC de l'ensemble d'essai sont très similaires entre modèles, les modèles de régression logistique offrant une performance seulement légèrement supérieure. La valeur absolue de l'ASC est de 0,71 pour la dépression et varie entre 0,59 et 0,63 pour l'anxiété, ce qui semble indiquer une capacité discriminatoire médiocre. Comme pour les résultats continus, les performances des ensembles d'entraînement et d'essai sont semblables, laissant entrevoir un surajustement minimal.

Importance relative des variables

Les pourcentages d'importance relative des variables pour les résultats continus (CESD-10, GAD-7, FS) sont présentés sur la figure 2. Pour les résultats liés aux échelles CESD-10 et GAD-7, les modèles CART, CTREE et de régression logistique ont tous systématiquement classé parmi les trois variables les plus importantes le sentiment d'appartenance à l'école, le fait

FIGURE 1
Arbre CART pour déterminer la présence de symptômes d'anxiété pertinents sur le plan clinique (GAD-7 ≥ 10)



Abbreviations : CART, arbre de classification et de régression; GAD-7, échelle d'évaluation du trouble d'anxiété généralisée (Generalized Anxiety Disorder); Sent. appart., sentiment d'appartenance à l'école.

Remarque : n désigne le nombre d'élèves dans le sous-groupe; p désigne le pourcentage de sujets dans le sous-groupe qui présentent des symptômes d'anxiété pertinents sur le plan clinique.

d'avoir une vie familiale heureuse et le sexe. Tous les modèles ont également classé la variable du temps de sommeil au quatrième rang pour ce qui est de l'importance relative, sauf le modèle de régression linéaire de l'anxiété qui a classé l'intimidation au quatrième rang. Cependant, les modèles CART et CTREE ont accordé plus de poids aux variables les mieux classées par rapport au modèle de régression linéaire. Dans les modèles CART et CTREE, les quatre principales variables comptent pour 78 % à 87 % de l'importance totale, tandis que dans le modèle de régression linéaire, elles comptent pour seulement 47 % de l'importance totale, le reste étant réparti également entre les autres variables du modèle.

Des résultats semblables sont observables pour l'échelle d'évaluation de l'épanouissement (FS), et bien que le sexe n'ait été jugé important dans aucun des modèles. Le fait de se sentir capable de parler de ses problèmes avec des ami(e)s a été classé parmi les quatre principales variables, la famille a été classée comme étant

importante dans les modèles CART et CTREE, et l'argent de poche a été classé comme étant important dans le modèle de régression linéaire. Là aussi, dans les modèles CART et CTREE, les quatre principales variables comptent pour 86 % à 93 % de l'importance totale, tandis que dans le modèle de régression linéaire, elles comptent pour seulement 43 % de l'importance totale.

Les pourcentages d'importance relative des variables pour les résultats binaires sont présentés sur la figure 2. Comme pour les résultats continus, le sentiment d'appartenance à l'école, le fait d'avoir une vie familiale heureuse et le sexe ont tous été systématiquement classés parmi les trois variables les plus importantes dans les modèles relatifs à la dépression et à l'anxiété. Tous les modèles relatifs à la dépression ont également classé le fait de pouvoir parler de ses problèmes avec sa famille au quatrième rang, tandis que le fait d'avoir été victime d'intimidation a été classé au quatrième rang dans tous les modèles relatifs à l'anxiété. Les quatre

principales variables comptent pour 92 % à 93 % de l'importance totale dans le modèle CART, tandis qu'elles comptent pour 79 % à 83 % de l'importance totale dans le modèle CTREE et pour 44 % à 46 % de l'importance totale dans le modèle de régression logistique.

Analyse

Cette étude fournit un aperçu de la méthodologie utilisée pour deux types d'arbres décisionnels, les arbres CART et les arbres CTREE, et les compare aux méthodes classiques de régression linéaire et logistique à l'aide d'une nouvelle application aux données d'enquête à grande échelle sur la santé mentale des jeunes. Elle vient enrichir les données probantes limitées dont on dispose sur la performance des arbres décisionnels dans ce domaine^{17,21,23,27}, par le biais de l'analyse d'un vaste échantillon de jeunes et d'un large éventail de prédicteurs. L'étude porte également sur les facteurs relatifs à la méthodologie des arbres décisionnels dans le contexte de la recherche en surveillance de la population,

TABEAU 2
Modèle de régression logistique pour déterminer la probabilité de présenter des symptômes d'anxiété pertinents sur le plan clinique (GAD-7 ≥ 10)

Variable	Valeur	RCA (95% CI)
Sexe (réf. = féminin)	Masculin	0,33 (0,31 à 0,34)***
Âge (en années)	Par année	1,05 (1,02 à 1,07)***
Origine ethnique (réf. = blanche)	Noire	0,5 (0,43 à 0,59)***
	Asiatique	0,73 (0,66 à 0,81)***
	Latino-américaine	0,83 (0,7 à 0,98)*
Argent de poche (réf. = 0 \$)	Autre/mixte	1,01 (0,94 à 1,09)
	1 \$ à 20 \$	0,93 (0,85 à 1,01)
	21 \$ à 40 \$	0,86 (0,77 à 0,95)**
	41 \$ à 100 \$	0,87 (0,79 à 0,96)**
	Plus de 100 \$	0,94 (0,86 à 1,03)
Province (réf. = Alberta)	Ne sait pas	0,87 (0,79 à 0,96)**
	Colombie-Britannique	0,89 (0,77 à 1,03)
	Ontario	0,92 (0,81 à 1,05)
Indice d'urbanisation (réf. = grand centre urbain)	Québec	0,66 (0,58 à 0,76)***
	Centre urbain de taille moyenne	1,02 (0,93 à 1,12)
Perception du poids (réf. = poids insuffisant)	Petit centre urbain/milieu rural	0,86 (0,80 à 0,91)***
	Poids plutôt idéal	0,78 (0,72 à 0,84)***
Prend un déjeuner chaque jour	Surpoids	1,03 (0,95 à 1,12)
	Oui	0,76 (0,72 à 0,8)***
Portions de fruits et de légumes	Nombre de portions	1,03 (1,01 à 1,04)***
Temps d'écran (heures)	Nombre d'heures	1,05 (1,05 à 1,05)***
Temps de sommeil (heures)	Nombre d'heures	0,83 (0,83 à 0,83)***
Tabagisme	Oui	1,12 (1,00 à 1,25)*
Utilisation de cigarettes électroniques	Oui	1,08 (1,01 à 1,15)*
A été victime d'intimidation au cours des 30 derniers jours	Oui	2,03 (1,88 à 2,18)***
Envisage de faire des études postsecondaires	Oui	1,16 (1,09 à 1,24)***
	1 ou 2 cours	1,06 (0,99 à 1,13)
	3 à 5 cours	1,16 (1,06 à 1,28)**
Cours manqués au cours des 4 dernières semaines (réf = 0 cours)	6 cours ou plus	1,23 (1,10 à 1,39)***
	Score unitaire	0,88 (0,87 à 0,89)***
Vie familiale heureuse	Oui	0,50 (0,47 à 0,54)***
Parle de ses problèmes avec sa famille	Oui	0,73 (0,69 à 0,77)***
Parle de ses problèmes avec ses ami(e)s	Oui	0,75 (0,71 à 0,8)***

Abbreviations : IC, intervalle de confiance; RCA, rapport de cotes ajusté; réf., groupe de référence.

* $p < 0,05$

** $p < 0,01$

*** $p < 0,001$

où l'exactitude prédictive doit être évaluée en fonction de la facilité d'interprétation du modèle. Au-delà des connaissances sur le sujet tirées des résultats de cette application à la santé mentale des jeunes, les répercussions analysées ci-dessous peuvent servir de guide aux chercheurs qui analysent

d'autres ensembles de données d'enquêtes à grande échelle.

Pour ce qui est de l'exactitude prédictive en matière de résultats à échelle linéaire, la régression linéaire a surpassé les méthodes CART et CTREE, affichant des valeurs R^2_{aj}

de 4 % à 5 % plus élevées et des EMQ de 3 % à 5 % moins élevés. Le nombre de paramètres du modèle s'est révélé similaire dans le cas de la méthode CART et de la régression linéaire, alors que la méthode CTREE a fourni des modèles plus complexes. Cependant, alors que la méthode CART et la régression linéaire ont compté un nombre similaire de paramètres, la méthode CART a permis de cerner beaucoup moins de variables uniques significatives, le nombre élevé de paramètres étant plutôt attribuable à des divisions multiples sur les mêmes variables prédictives continues. À l'inverse, les modèles de régression ont prédit un effet linéaire des variables continues et ont fourni une seule estimation, correspondant à l'effet d'une augmentation d'une unité de la variable quelle que soit la valeur initiale.

En ce qui concerne les résultats binaires, les modèles de régression logistique ont présenté là aussi une performance prédictive supérieure aux modèles CART et CTREE, mais cette performance globale s'est révélée plus proche que pour les résultats continus, avec une exactitude prédictive de 1 % à 2 % supérieure et une ASC de 0 % à 3 % supérieure. La méthode CART a généré des modèles beaucoup plus parcimonieux que la méthode CTREE et la régression logistique, tant en ce qui concerne le nombre total de paramètres que le nombre de variables uniques. Des études antérieures de petite envergure réalisées par Burke et ses collaborateurs²¹, Mitsui et ses collaborateurs¹⁵ et Handley et ses collaborateurs²⁷ ont montré que la valeur de l'ASC était de 4 % à 8 % inférieure avec la méthode CART comparativement à la régression logistique, tandis que, à l'inverse, une étude de Batterham et ses collaborateurs¹⁷ a montré que l'ASC était de 2 % supérieure avec la méthode CART par rapport à la régression logistique. Bien qu'il soit difficile d'effectuer une comparaison directe des résultats de l'ASC issus de ces études en raison des différences entre échantillons, résultats et spécifications des modèles, il convient tout de même de noter que la performance obtenue dans les études avec les deux techniques ne diffère pas considérablement.

Ainsi, bien que les analyses de régression linéaire et logistique puissent fournir de légers avantages sur le plan de la valeur prédictive, la simplicité des modèles générés au moyen de la méthode CART pourrait être préférable, en particulier pour l'application des connaissances dans le

TABLEAU 3

Comparaison de l'exactitude prédictive pour les résultats continus et les résultats binaires des modèles CART, CTREE et de régression

Résultats continus	Méthode	N ^{bre} de paramètres	N ^{bre} de variables uniques	Données d'entraînement R^2_{aj}	Données d'entraînement EMQ	Données d'essai R^2_{aj}	Données d'essai EMQ
CESD-10	CART	38	9	0,35	4,73	0,33	4,76
	CTREE	57	10	0,36	4,70	0,34	4,73
	Régression linéaire	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
	Régression linéaire	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
	Régression linéaire	40	24	0,51	3,79	0,51	3,78
Résultats binaires	Méthode	N ^{bre} de paramètres	N ^{bre} de variables uniques	Données d'entraînement tEC	Données d'entraînement ASC	Données d'essai tEC	Données d'essai ASC
Dépression	CART	9	6	0,75	0,71	0,74	0,70
	CTREE	53	14	0,75	0,71	0,74	0,70
	Régression logistique	39	22	0,76	0,71	0,76	0,70
Anxiété	CART	8	5	0,80	0,60	0,79	0,59
	CTREE	52	11	0,80	0,61	0,79	0,61
	Régression logistique	34	20	0,80	0,63	0,80	0,63

Abbreviations : ASC, aire sous la courbe caractéristique de la performance d'un test; CART, arbre de classification et de régression; CESD-10, échelle d'évaluation de la dépression à 10 items du Center for Epidemiologic Studies; CTREE, arbre d'inférence conditionnelle; EMQ, écart moyen quadratique; FS, échelle d'évaluation de l'épanouissement (Flourishing Scale); GAD-7, échelle d'évaluation du trouble d'anxiété généralisée à 7 items (Generalized Anxiety Disorder); R^2_{aj} , valeur ajustée de R^2 ; tEC, taux d'exactitude de la classification.

contexte de la recherche en santé de la population, où l'accent est mis sur la compréhension des liens et la communication des résultats à un public non spécialisé.

Les modèles d'arbres décisionnels et les modèles de régression ont cerné de manière systématique les mêmes ensembles de prédicteurs importants pour chaque résultat, ce qui témoigne d'un niveau général de consensus entre les méthodes. Cependant, les modèles CART et CTREE ont accordé une importance relative beaucoup plus grande aux principaux prédicteurs par rapport aux modèles de régression, attribuant plus de 75 % de l'importance totale aux quatre principaux prédicteurs, comparativement aux modèles de régression, qui ont attribué quant à eux moins de 50 % de l'importance totale aux principaux prédicteurs. Cela concorde avec la parcimonie supérieure observée dans les modèles CART et CTREE et cela met en évidence la capacité des arbres décisionnels à repérer les facteurs les plus importants.

De plus, une limite courante des modèles de régression est que les facteurs présentant

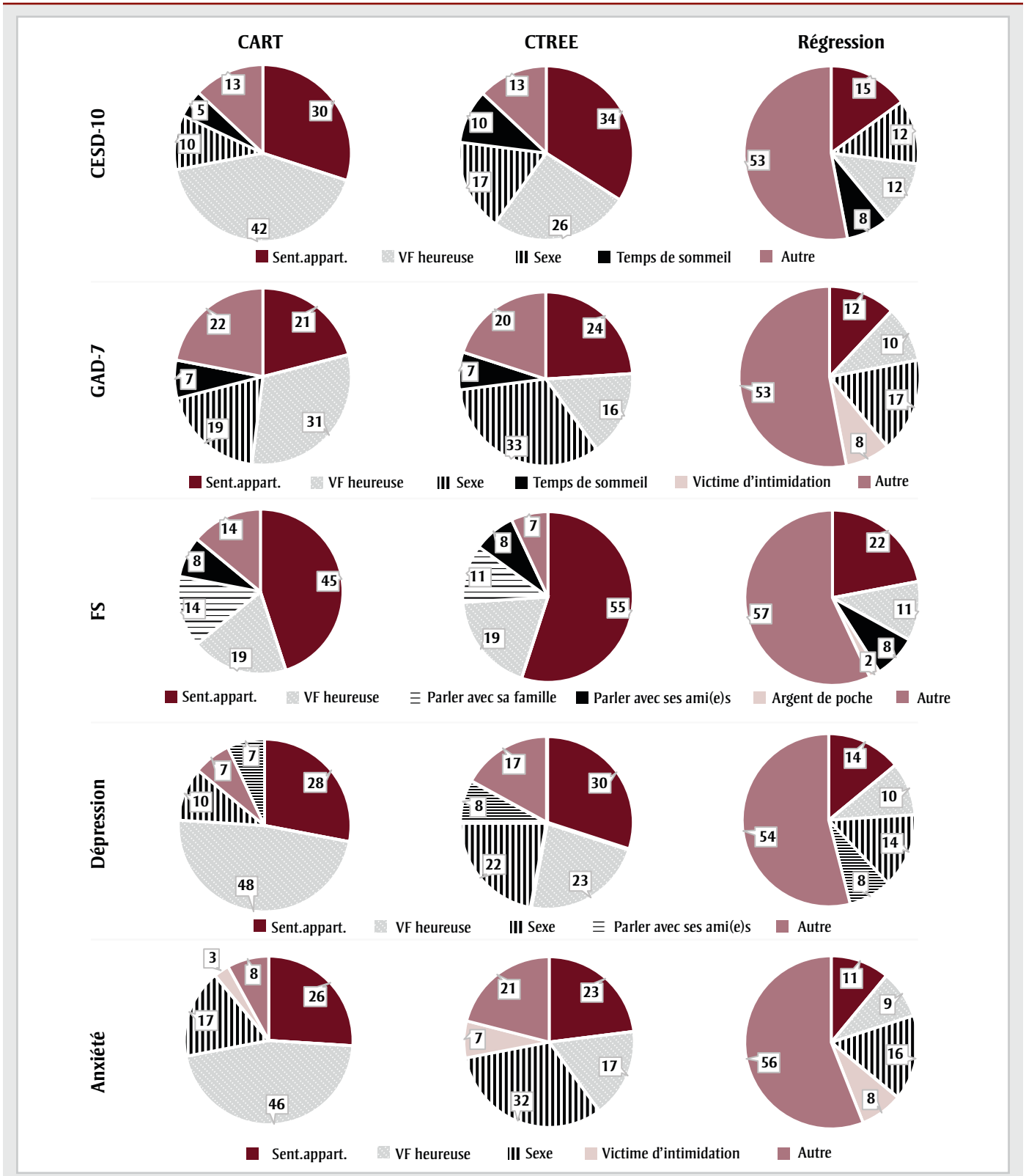
une multicollinéarité élevée ont tendance à être « éliminés » lorsqu'ils sont saisis simultanément, ce qui entraîne une surestimation de la variance ou un biais d'omission de variable pouvant conduire à ce que des facteurs ne soient pas pris en compte⁴⁵. Ce phénomène a été observé dans des recherches antérieures comparant les arbres et l'analyse de régression²³, ce qui semble indiquer que les méthodes d'arbres décisionnels peuvent offrir une représentation plus claire des facteurs clés pour faciliter la prise de décisions. Cet avantage en matière de parcimonie peut s'avérer particulièrement utile dans le domaine de la recherche en prévention des maladies à l'échelle de la population, où une multitude de facteurs de risque et de facteurs de confusion concurrents peuvent être présents.

Un plus grand sentiment d'appartenance à l'école et une vie familiale heureuse ont été systématiquement identifiés comme des prédicteurs clés et ont été associés à des niveaux inférieurs de dépression et d'anxiété et à des niveaux supérieurs d'épanouissement. Cette constatation concorde avec les résultats d'études antérieures

qui établissent un lien entre, d'une part, les relations familiales et l'anxiété chez les adolescents¹¹ et, d'autre part, le sentiment d'appartenance à l'école et la détresse émotionnelle et la dépression chez les jeunes^{8,9}. De plus, des analyses antérieures d'arbres de classification portant sur des adolescentes ont révélé qu'un fonctionnement scolaire déficient constituait un facteur de risque majeur de dépression et que le soutien parental n'offrait une protection que chez les sous-groupes présentant de faibles symptômes de dépression au début de l'étude¹⁸. Le lien protecteur conféré par le sentiment d'appartenance à l'école met en évidence le rôle des milieux scolaires dans le façonnement de la santé mentale des jeunes et montre pourquoi les écoles constituent un milieu approprié d'intervention, compte tenu de leur capacité à rejoindre une grande partie de la population de jeunes. La méthodologie de l'arbre décisionnel décrite dans cette étude est bien adaptée aux études qui voudraient évaluer des caractéristiques environnementales complexes et des interventions réalisées simultanément.

Comme nous l'avons mentionné précédemment, l'un des avantages des arbres

FIGURE 2
 Pourcentages d'importance relative des principales variables prédictives des modèles CART, CTREE et de régression pour les résultats continus et les résultats binaires



Abréviations : CART, arbre de classification et de régression; CESD-10, échelle d'évaluation de la dépression à 10 items du Center for Epidemiologic Studies; CTREE, arbre d'inférence conditionnelle; GAD-7, échelle d'évaluation du trouble d'anxiété généralisée à 7 items (Generalized Anxiety Disorder); FS, échelle d'évaluation de l'épanouissement (Flourishing Scale); Sent. appart., sentiment d'appartenance à l'école; VF, vie familiale.

décisionnels est leur capacité à prendre en compte des interactions complexes entre prédicteurs et à cerner les sous-groupes à risque élevé qu'il convient de cibler dans le cadre des efforts de prévention et d'intervention. Dans l'exemple illustratif sur l'anxiété, l'intimidation a été associée de façon significative à la probabilité de présenter des symptômes d'anxiété pertinents sur le plan clinique dans le modèle de régression, alors que dans le modèle CART, l'intimidation apparaît seulement comme un facteur de risque de présenter des niveaux plus élevés d'anxiété chez le sous-ensemble d'élèves de sexe féminin ayant une vie familiale heureuse et un faible sentiment d'appartenance à l'école.

De même, le temps de sommeil a été associé à des probabilités plus élevées d'anxiété dans le modèle de régression, quoique l'effet soit de faible ampleur, alors qu'à l'inverse, le modèle CART a montré que le sommeil était un facteur de protection chez les élèves de sexe féminin n'ayant pas une vie familiale heureuse et n'ayant pas un sentiment d'appartenance à l'école élevé. Les estimations issues du modèle de régression correspondent à l'association globale moyenne observée pour l'ensemble de l'échantillon et ne fournissent pas d'indication sur les répercussions différentielles dans divers sous-groupes. Dans le cas présent, la faible ampleur de l'effet en ce qui concerne le temps de sommeil dans le modèle de régression masque l'importance de ce facteur au sein d'un sous-groupe précis.

Des études réalisées par Handley et ses collaborateurs²⁷ et Batterham et ses collaborateurs²⁸, qui ont porté sur les idées suicidaires chez les adultes, ont chacune fait état de facteurs importants présents dans les analyses d'arbres décisionnels qui ne s'étaient pas révélés significatifs dans les modèles de régression correspondants. Comme le mentionne Handley, cela indique la présence d'un impact multiplicatif plutôt qu'indépendant pour ces facteurs, lesquels ne seraient pas détectés par un modèle de régression standard sur les effets principaux. Ainsi, les arbres décisionnels peuvent être beaucoup plus utiles que les modèles de régression pour les chercheurs et les praticiens qui cherchent à cerner les caractéristiques spécifiques des groupes les plus à risque afin d'adapter les interventions qui leur sont destinées.

Malgré ces constatations, la meilleure performance prédictive des modèles de régression comparativement aux modèles d'arbres décisionnels observée dans notre étude pourrait donner à penser que la nature sous-jacente des prédicteurs est relativement linéaire. Dans l'exemple portant sur l'anxiété, nous avons pu montrer que le sentiment d'appartenance à l'école était un facteur important dans les deux sous-groupes d'élèves ayant une vie familiale heureuse, alors que le sexe s'est révélé être le facteur important suivant dans trois des quatre sous-groupes subséquents. Cela donne à penser que l'effet de ces facteurs est similaire à l'échelle de l'ensemble de l'échantillon, et donc qu'une analyse de régression permettrait de saisir adéquatement cet effet au moyen d'une seule estimation par le modèle. Les arbres décisionnels offrent un avantage plus important par rapport aux modèles de régression lorsque les véritables relations sous-jacentes entre les données sont non linéaires¹². Les chercheurs devraient donc analyser attentivement les structures de données sous-jacentes, par une exploration théorique et descriptive, lorsqu'ils cherchent à sélectionner la technique d'analyse la plus appropriée.

Cette étude a porté sur deux types d'arbres de décision : CART et CTREE. Ces deux modèles segmentent la population en sous-groupes grâce à la sélection récursive des variables et des seuils qui génèrent une séparation maximale entre les sous-groupes et une variabilité minimale au sein d'un même groupe. Bien que les méthodes CART et CTREE aient offert une performance similaire quant à l'exactitude prédictive, la méthode CART a systématiquement généré des modèles plus parcimonieux, et notamment un nombre total inférieur de paramètres de modèles et de variables uniques. Les modèles CART et CTREE ont généralement présenté des divisions multiples pour différentes valeurs de la même variable, en particulier dans le cas des résultats continus étudiés. La tendance à privilégier les prédicteurs continus par rapport aux prédicteurs catégoriels en raison du nombre élevé de divisions possibles est l'un des inconvénients fréquemment signalés en ce qui concerne les arbres décisionnels^{12,31}. Dans le cas des résultats binaires, cette limite semble être plus préoccupante pour les modèles CTREE que pour les modèles CART.

Un autre inconvénient souvent évoqué relativement aux arbres décisionnels est la

tendance des modèles à se surajuster aux données de l'échantillon³⁵, un phénomène qui est en partie atténué par l'élagage dans le cas du modèle CART et par les critères d'arrêt fondés sur les tests de signification statistique dans le cas du modèle CTREE³⁵. Dans notre étude, la performance similaire des modèles avec l'ensemble d'entraînement et l'ensemble d'essai montre que le surajustement n'est pas un problème pour l'une ou l'autre de ces méthodes, ce qui pourrait être attribuable à la taille importante de l'échantillon dans cet ensemble de données. Fait intéressant, la méthode CTREE a généré des modèles beaucoup plus complexes que la méthode CART. Les modèles CTREE présentés dans l'étude ont fonctionné avec un seuil de signification statistique standard de $\alpha = 0,05$ avec correction de Bonferroni, ce qui laisse supposer que des critères plus rigoureux devraient sans doute être utilisés avec les modèles CTREE pour les échantillons de grande taille. Ainsi, même si la littérature privilégie généralement les modèles CTREE¹², notre étude laisse penser que les chercheurs travaillant avec des données en santé à grande échelle devraient plutôt envisager d'utiliser les modèles CART lorsque leurs principales préoccupations portent sur la parcimonie et sur la facilité d'interprétation du modèle.

Forces et limites

Cette étude offre une nouvelle application des arbres décisionnels aux données d'enquêtes canadiennes en santé à grande échelle. Contrairement aux études précédentes, limitées, cette étude tire profit d'un échantillon de grande taille qui permet d'élaborer des structures arborescentes complexes.

Pendant, la complexité résultante des arbres rend l'interprétation difficile, ce qui a pour effet de réduire l'un des principaux avantages de l'analyse des arbres. Bien que l'étude utilise des critères d'arrêt et d'élagage standard, des restrictions supplémentaires comme la limitation du nombre de niveaux et l'utilisation de seuils de signification pourraient générer des arbres de plus petite taille, plus faciles à interpréter. Les répercussions de diverses restrictions sur l'ajustement global du modèle devraient être prises en compte dans des travaux futurs. De plus, comme seuls les effets principaux ont été pris en compte dans les modèles de régression analysés dans notre étude, l'inclusion de paramètres d'interaction aurait pu améliorer

la performance relative, mais, comme nous l'avons mentionné précédemment, cela est susceptible d'entraîner des problèmes de calcul et d'interprétation.

Le faible ajustement global du modèle constitue une autre limite de l'étude. Les valeurs R^2_{aj} de l'ensemble d'essai pour les résultats continus variaient entre 0,27 et 0,51, ce qui signifie que les prédicteurs pris en compte expliquent moins de la moitié des variations globales observées dans les résultats. L'ASC pour les résultats binaires variait entre 0,59 et 0,70, ce qui indique une capacité discriminative faible à modérée. Bien qu'il ne soit pas rare que les études comportementales présentent un faible ajustement de modèle, ce résultat donne à penser que d'autres facteurs intrinsèques non visés par notre étude pourraient jouer un rôle important dans la prédiction des résultats en matière de santé mentale. Dans des études antérieures sur les résultats liés aux idées suicidaires, l'ASC était plus élevée, se situant à environ 0,80^{15,21,27}, mais ces études intégraient la présence de dépression au début de l'étude, ce qui constitue un facteur de prédiction déjà bien établi.

De plus, notre étude est transversale et non randomisée, de sorte que ni les arbres décisionnels ni les modèles de régression ne peuvent faire la preuve de relations de cause à effet entre les prédicteurs et les résultats en matière de santé mentale. De manière plus globale, les arbres décisionnels sont généralement considérés comme des méthodes exploratoires¹² utilisées pour la formulation d'hypothèses. De plus, les arbres décisionnels ne sont pas des méthodes déterministes et sont très sensibles à l'échantillon et au choix des paramètres. Des méthodes comme les forêts aléatoires, qui génèrent la croissance d'arbres de décision multiples et regroupent les résultats en mesures générales d'importance variable, ont été élaborées pour surmonter cette instabilité⁴⁶, mais au détriment de la facilité d'interprétation. Enfin, les méthodes CART et CTREE utilisées dans notre étude ne tiennent pas compte de la nature hiérarchique des données (c'est-à-dire du fait que les élèves sont regroupés au sein des écoles). De nouvelles méthodes, notamment RE-EM^{47,48} et M-CART⁴⁹, ont été élaborées pour prendre en compte cette absence d'indépendance des observations et seraient à utiliser dans le cadre de recherches futures.

Conclusion

Malgré leur utilisation croissante dans d'autres domaines, les arbres décisionnels demeurent une technique d'analyse sous-utilisée dans la recherche en santé de la population. Bien que la performance prédictive des arbres décisionnels soit légèrement inférieure à celle des méthodes de régression classiques, les arbres décisionnels offrent un moyen d'étudier les relations complexes entre prédicteurs, et présentent les résultats sous une forme qui est facile à interpréter par des publics non spécialisés, ce qui facilite l'application des connaissances. La capacité des arbres décisionnels à cerner les sous-groupes à risque élevé qu'il convient de cibler dans le cadre des efforts de prévention et d'intervention est particulièrement utile pour les praticiens en santé publique qui disposent de ressources limitées. Les arbres décisionnels peuvent constituer un puissant outil à ajouter au répertoire des méthodes des chercheurs en santé de la population pour répondre aux questions de recherche auxquelles les méthodes de régression classiques ne peuvent pas répondre.

Remerciements

L'étude COMPASS a reçu le soutien d'une subvention transitoire de l'Institut de la nutrition, du métabolisme et du diabète des Instituts de recherche en santé du Canada (IRSC), grâce à l'attribution du financement prioritaire « Obesity Interventions to Prevent or Treat » (Interventions pour prévenir ou traiter l'obésité) (OOP-110788; subvention accordée à SL), d'une subvention de fonctionnement de l'Institut de la santé publique et des populations des IRSC (MOP-114875; subvention accordée à SL), d'une subvention de projet des IRSC (PJT-148562; subvention accordée à SL), d'une subvention transitoire des IRSC (PJT-149092; subvention accordée à KP et SL), d'une subvention de projet des IRSC (PJT-159693; subvention accordée à KP), d'un accord de financement de la recherche conclu avec Santé Canada (n° 1617-HQ-000012; contrat attribué à SL) et d'une subvention d'équipe IRSC-Centre canadien sur les dépendances et l'usage de substances (CCDUS) (OF7 B1-PCPEGT 410-10-9633; subvention accordée à SL). L'étude COMPASS au Québec est également financée par le ministère de la Santé et des Services sociaux du Québec et par la Direction régionale de santé publique du CIUSSS de la Capitale-Nationale. LD a reçu

du financement du Conseil de recherche en sciences naturelles et en génie du Canada (RGPIN-2016-04396).

Conflits d'intérêts

Scott Leatherdale est rédacteur scientifique adjoint de la revue PSPMC, mais il s'est retiré du processus d'évaluation de cet article. Les auteurs n'ont aucun conflit d'intérêts à déclarer.

Contributions des auteurs et avis

KB a conçu la recherche, effectué les analyses statistiques et rédigé le manuscrit dans le cadre de sa thèse de doctorat à l'Université de Waterloo. SL et JD ont supervisé KB pour la conception des travaux et la rédaction du manuscrit. LD a contribué à l'analyse et à l'interprétation des données. LD et KP ont proposé des idées et des pistes de réflexion pour enrichir la discussion et ont révisé le manuscrit du point de vue de son contenu intellectuel important. SL est le chercheur principal de l'étude COMPASS et a dirigé la mise en œuvre de l'étude. KP a conçu le module sur la santé mentale de l'étude COMPASS. Tous les auteurs ont orienté le plan d'analyse, ont formulé des commentaires sur les versions préliminaires et ont approuvé la version définitive du manuscrit.

Le contenu de l'article et les points de vue qui y sont exprimés n'engagent que les auteurs; ils ne correspondent pas nécessairement à ceux du gouvernement du Canada.

Références

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):742-754. <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):2097-2108. <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):1383-1410. <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;182:137-143. <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):278-293. <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):241-252. <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):221-230. <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):1061-1067. <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):723-737. <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):1253-1266. <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):172-181. 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):444-453. <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;262:175-183. <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;269:474-480. <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):702-711. <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;11:2367-2375. <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):1325-1335. <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):306-314. <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):329-348. <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.
32. Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating.* New York (NY): Springer; 2009.
33. James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning.* New York (NY): Springer; 2013.
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):651-674. <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):128-132. <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):408-412. <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):1092-1097. <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):227-234A.
41. Hone L, Jarden A, Schofield G. Psychometric properties of the flourishing scale in a New Zealand sample. *Soc Indic Res.* 2014;119(2):1031-1045. <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):903-915. <https://doi.org/10.1007/s11205-014-0663-1>
43. Statistique Canada. GéoRecherche [Internet]. Ottawa (Ont.): Statistique Canada; 2017 [consultation le 2 octobre 2019]. En ligne à : <https://www12.statcan.gc.ca/census-recensement/2016/geo/geosearch-georecherche/index-fra.cfm?LANG=F>
44. Statistique Canada. Télécharger, Profil du recensement, Recensement de 2016 [Internet]. Ottawa (Ont.) : Statistique Canada; 2017 [consultation le 2 octobre 2019]. En ligne à : https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/download-telecharger/comp/page_dl-tc.cfm?Lang=F
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):667-677. <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):323-348. <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):578-592. <https://doi.org/10.1080/00273171.2018.1552555>