Administration and Policy in Mental Health and Mental Health Services Research
https://doi.org/10.1007/s10488-020-01045-y

ORIGINAL PAPER

Leveraging Machine Learning to Identify Predictors of Receiving Psychosocial Treatment for Attention Deficit/Hyperactivity Disorder

Anne S. Morrow[1,2,3] · Alexandro D. Campos Vega[4] · Xin Zhao[1,2] · Michelle M. Liriano[1]

Abstract

This study aimed to identify factors associated with receiving psychosocial treatment for ADHD in a nationally representative sample. Participants were 6630 youth with a parent-reported diagnosis of ADHD from the 2016–2017 National Survey of Children's Health. Machine learning analyses were performed to identify factors associated with receipt of psychosocial treatment for ADHD. We examined potentially associated factors in the broad categories of variables hypothesized to affect problem recognition (e.g., severity, mental health comorbidities); the decision to seek treatment; service selection (e.g., insurance coverage) and service use. We found that three machine learning models unanimously identified parent-reported ADHD severity (mild vs. moderate/severe) as the factor that best distinguishes between children who receive psychosocial treatment for ADHD and those who do not. Receive operating characteristic curve analysis revealed the following model performance: classification and regression tree analysis (area under the curve; AUC = .68); an ensemble model (AUC = .71); and a deep, multi-layer neural network (AUC = .72), as well as comparison to a logistic regression model (AUC = .69). Further, insurance coverage of mental/behavioral health needs emerged as a salient factor associated with the receipt of psychosocial treatment. Machine learning models identified risk and protective factors that predicted the receipt of psychosocial treatment for ADHD, such as ADHD severity and health insurance coverage.

Keywords ADHD · National sample · Machine learning · Predictors · Psychosocial treatment · Health insurance coverage

Attention-deficit/Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder that affects nearly one in ten of children in the United States (Danielson et al. 2018). ADHD is not only common, it is costly: in addition to treatment expenses, raising a child with ADHD costs families five times more than a typically developing child (e.g., parents of children with ADHD were more likely to have been fired; Zhao et al. 2019). Notably, a childhood history of ADHD increases risk for a host of impairments throughout adolescence and adulthood, from alcohol/substance abuse, to poorer educational attainment, as well as financial instability (Altszuler et al. 2015; Kuriyan et al. 2013; Merrill et al. 2019). Given the impairing nature of ADHD, several evidence-based treatments for ADHD have been developed: psychosocial treatment, medication treatment, and the combination of these two treatment modalities (Evans et al. 2018; Jensen et al. 2001). Despite these promising options to intervene in the lives of children with ADHD, there is a gap between those who need treatment and those who receive evidence-based treatment (DuPaul et al. 2018). For example, less than a third of students with ADHD receive psychosocial treatment at school, and certain students are less likely to receive psychosocial treatment than others (e.g., female vs. male students with ADHD; DuPaul et al. 2018). However, research on the correlates of receiving psychosocial treatment for ADHD has several limitations, such as a reliance on treatment-seeking samples (Murray et al. 2014; Waxmonsky et al. 2019; Zendarski et al. 2018). The primary

The following paper was presented (preliminary analyses) at the Association for Behavioral and Cognitive Therapy meeting, November 2018. Submitted 6/1/2019.

✉ Anne S. Morrow
amorr036@fiu.edu

1 Department of Psychology, Florida International University, Miami, USA

2 Center for Children and Families, Florida International University, Miami, USA

3 College of Psychology, Nova Southeastern University, Davie, USA

4 Zyanya Tech, LLC, Miami, USA

Published online: 13 May 2020

aim of the current study is to identify predictors of receiving psychosocial treatment for ADHD among youth from a large, nationally drawn sample.

More research on the science-to-practice gap in the field of ADHD treatment would be helpful regarding several of the evidence-based interventions (psychosocial treatment, medication treatment, and their combination; Evans et al. 2018; Jensen et al. 2001). However, we elected to focus on identifying predictors of psychosocial treatment (vs. medication) for ADHD for several reasons: (1) current treatment guidelines, (2) parent preferences, and (3) cost. First, we clarify that behavior therapy is recommended as the first-line of treatment for children under the age of six (American Academy of Pediatrics, Subcommittee on ADHD 2011). For children six years and older, guidelines recommend either medication or behavior therapy, yet "preferably both" would be prescribed (AAP 2011). In other words, the American Academy of Pediatrics (2011) guidelines recommend that children with ADHD of all ages would "preferably" receive psychosocial treatment. Interestingly, despite the recommendation to receive both treatments, a recent report found 45% of Medicaid-enrolled children treated for ADHD did not receive any behavior therapy when prescribed medication (Office of the Inspector General, OIG; 2019). Further, Waschbusch et al. (2011) found that a majority of parents prefer psychosocial treatment to medication treatment. Although the literature on ADHD treatment preferences is somewhat mixed (Schatz et al. 2015), there is growing, yet nuanced evidence on parents' negative attitudes toward stimulant medications (e.g., Ng et al. 2017). Lastly, Pelham et al. (2016) found that a sequential, multiple assignment, randomized trial revealed that beginning with psychosocial treatment for ADHD (vs. medication) led to lower rates of observed classroom rule violations during the school year. Importantly, parents who were assigned to start treatment with behavioral parent training exhibited higher rates of attendance than those who began with medication first. Over the course of a school year of treatment, beginning with behavioral treatment also cost less money than beginning with medication (Page et al. 2016). In sum, there are several evidence-based treatment options for children with ADHD (Evans et al. 2018; Jensen et al. 2001), and we focused on psychosocial treatment in the current study because of real-world considerations such as expert guidelines, patient preferences, and cost.

Predictors of Receiving Psychosocial Treatment for ADHD: Limitations of the Current Literature

Theoretically associated factors with receiving treatment for ADHD are detailed and span four stages from problem recognition, to the decision to seek help, service selection, and service use (i.e., the ADHD helping-seeking behavior model; Eiraldi et al. 2006). Research on these factors associated with receiving psychosocial treatment for ADHD has several salient limitations. There is a reliance on data from treatment-seeking youth (e.g., Murray et al. 2014), as well as youth recruited via clinical settings such as pediatric practices (e.g., Zendarski et al. 2018). Studies with samples collected in clinical settings identified correlates of receiving psychosocial treatment for ADHD such as: greater academic/behavioral needs (falling within the problem recognition stage), attending private (Catholic) school (falling within the service selection stage), and having a formal education plan (falling within the service use characteristics stage; Murray et al. 2014; Zendarski et al. 2018). Interestingly, there was also a recent retrospective study, a review of commercial insurance claims, among more than 800,000 youth receiving healthcare services for ADHD (Waxmonsky et al. 2019). Significant correlates of receiving behavior therapy for ADHD were consistent with the help-seeking behavior model (Eiraldi et al. 2006): ADHD severity and comorbid behavioral disorders (hypothesized to affect problem recognition); gender, region of residence, psychosocial therapy use by siblings (hypothesized to possibly affect the decision to seek help); as well as receiving medication treatment for ADHD (a service use characteristic; Waxmonsky et al. 2019). To our knowledge, there are only two previous studies analyzing the correlates of receipt of psychosocial treatment for ADHD among nationally drawn samples (DuPaul et al. 2018; Zablotsky et al. 2018). These studies identified factors associated with receiving psychosocial treatment for ADHD such as comorbidity with other mental health problems (hypothesized to affect problem recognition; Zablotsky et al. 2018), as well as hyperactive-impulsive symptom presentation, ADHD severity (hypothesized to affect problem recognition); male gender, and lower family income (hypothesized to affect the decision to seek help); and whether or not the child was covered by health insurance ((hypothesized to affect service selection; DuPaul et al. 2018). In sum, the empirical research on predictors of receiving psychosocial treatment for ADHD buttresses Eiraldi et al. (2006) theoretical conceptualization. However, several gaps remain in the literature, such as limited studies with nationally drawn samples.

Machine Learning Techniques may Improve Predictive Analytics

Perhaps we should look to education researchers for an example of success in identifying predictors of youth outcomes: machine learning techniques have demonstrated particular promise to help close the research-to-practice gap by forecasting high school dropout (e.g., Sara et al. 2015; Aguiar et al. 2015; Knowles 2015; Sansone 2019). We define machine learning as a type of artificial intelligence that enables machines (algorithms) to improve with experience; that is, machine learning encompasses a field of study in which computers can learn without being explicitly programmed (Samuel 1959). To facilitate readability of our manuscript, we have included a brief glossary of common terms relevant to the machine learning literature (Table 1). In one example of applied machine learning research, the state of Wisconsin currently has deployed an early warning system in use with 225,000 students, leveraging a predictive model of high school dropout risk to provide accurate predictions (accuracy = 93. 5%, area under the curve = 0.96) without collecting any data beyond typical administrative records (Knowles 2015). Machine learning has also been used to predict dropout among university students (e.g., Aulck et al. 2016). Administrative datasets from the field of education are particularly ripe for data mining and predictive analytics; datasets often have large sample sizes and would likely be collected regardless of their utility for prediction. In particular, Receiver Operating Characteristic (ROC) curves have also been helpful to evaluate predictive models of risk factors for high school dropout (Bowers and Zhou 2019). For example, Bowers and Zhou (2019)

Table 1 Brief glossary of common machine learning terms

General terms	
Artificial intelligence (AI)	Techniques enabling computers to mimic human intelligence (e.g., machine learning, rule-based, etc.)
Machine learning (ML)	Type of artificial intelligence that enables machines to improve with experience
Deep learning	Type of machine learning that trains algorithms with vast data to perform tasks
Ensemble model	A method: using multiple learning algorithms to improve performance (e.g., "bagging")
Unsupervised learning	Type of machine learning: algorithms infer patterns without labeled outcomes
Supervised learning	Type of machine learning: algorithms infer patterns with known, labeled outcomes
Reinforcement learning	Type of machine learning: algorithms train by trial & error in interactive "environment"
Natural Language Processing (NLP)	Type of artificial intelligence: computers "understand" language, uses computational linguistics
Types of models	
Logistic regression	Binary classification model; estimates parameters of a logistic function based on data
Classification and regression trees (CART)	Non-parametric decision tree learning technique, uses recursive partitioning
Random Forest	Uses Bootstrap-aggregating "Bagging" decision trees; ensemble learning that helps w/overfitting
Gradient[a] Boosted Trees	Also called Gradient Boosting Machine (GBM); ensemble learning method, helps with overfitting
LASSO	Least absolute shrinkage & selection operator; type of regression analysis, helps variable selection
Neural Networks	A set of algorithms, modeled after human brain; designed to recognize patterns
Support vector machine (SVM)	Non-probabilistic binary linear classifier: uses a data transformation, "kernel trick"
Naïve Bayes	Based on Bayes' conditional probability rule, useful for high-dimensional spaces (many predictors)
Model Evaluation Terms	
Sensitivity	True positive rate, also called recall
Specificity	True negative rate
Precision	Positive predictive value; proportion of positive predictions that are true positives
Accuracy	Percentage of correct predictions
Confusion matrix	Table with 2 rows & columns: false positives & negatives, true positives & negatives
Receiver Operating Characteristic (ROC)	Based on signal detection theory; graph of diagnostic ability for a binary classifier
Area under the (ROC) curve (AUC/AUROC)	Area enclosed under ROC curve; with classifiers ranging from perfect (= 1) to completely random (= .5)

[a]A gradient is multi-variable generalization of the calculus concept, a derivative (i.e., for a single-variable function, derivative = the slope of the tangent to the graph)

found that most dropout flags (risk indicators) have low specificity, longitudinal models provided the most accurate flags, and the best-performing cross-sectional flags were low/failing grades. Machine learning techniques such as ensemble models may be particularly useful to improve upon previous prediction research relying heavily on risk indicators identified by traditional statistical methods (Bowers and Zhou 2019).

Further, a theme uniting the success of the literature on school dropout prediction is the division of datasets into "train" and "test" subsections (Aguiar et al. 2015; Knowles 2015; Sansone 2019; Sara et al. 2015); that is, algorithms were developed based on a portion of a dataset, and evaluated on a different portion of each dataset that was set aside for the purpose of model evaluation (Breiman 2001). This process of partitioning datasets into train and test components was also helpful across a number of other scientific disciplines and use cases (e.g., ozone levels, speech recognition), improving prediction performance by forgoing a reliance on the assumption that data follow a certain distribution (Breiman 2001). In sum, there are two cultures of statistical modeling (Breiman 2001): (1) traditional statistical models evaluated by goodness-of-fit, and (2) using algorithmic models that treat the data mechanism as unknown, evaluated by predictive accuracy.

Within the field of mental health, there is also an emerging literature on predictive analytics (Kessler et al. 2019). For example, ROC curves have been recognized as particularly helpful to evaluate prediction model performance in the growing subfield of machine learning in mental health (Caye et al. 2019; Youngstrom 2014). A recent review on predictors of self-injurious thoughts and behaviors noted the utility of machine learning techniques to identify subgroups and to select variables not yet identified in the literature as relevant for building prediction models (Burke et al. 2019). In addition to improving model performance, machine learning techniques also replicated findings on well-established risk factors for self-harming behavior (Burke et al. 2019). Other successes include a support vector machine, a type of machine learning classifier, achieved an accuracy of 84.5% (area under the curve; AUC = .84) in predicting response to methylphenidate in a sample of 83 Korean youth with ADHD (Kim et al. 2015). Among a sample of 61 Swedish adolescents, four machine learning models accurately predicted response to internet-delivered cognitive behavioral treatment for Obsessive–Compulsive Disorder (e.g., least absolute shrinkage and selection operator; LASSO model, accuracy = 75%; Random Forest model, accuracy = 75%; Lenhard et al. 2018). Further, machine learning has also been applied to the analysis of language data in the context of analyzing therapist and patient speech, gaining insights about therapy via algorithms trained on manually coded utterances (Cummins et al. 2019; Huber et al. 2019; Pérez-Rosas et al. 2016, 2017). Lastly, an unblinded trial ($n = 70$) found that participants who interacted for two-weeks with an artificially intelligent "chat-bot" that delivered cognitive-behavioral therapy content self-reported significantly lower depression symptoms than an education control group (Fitzpatrick et al. 2017).

On the other hand, we note that there is promise for machine learning techniques to improve predictive analytics, yet some mixed findings: one study compared various machine learning techniques (e.g., LASSO) to traditional statistical methods to evaluate prediction of pediatric bipolar disorder diagnosis (Youngstrom et al. 2018). Less complex models that focused on predictors previously established by the literature outperformed more complex machine learning models, and notably these simpler a priori models generalized as well or better to new samples (Youngstrom et al. 2018).

What We Know and What We Don't Know

In sum, we know that the use of machine learning techniques to predict outcomes is growing within the field of mental health research (e.g., Burke et al. 2019). Other fields, such as education research, have successfully deployed accurate models in statewide systems (e.g., Knowles 2015). To our knowledge, mental health providers have not yet deployed machine learning prediction models in practice in systems of care. In addition, several predictors of receiving psychosocial treatment for ADHD have been identified in the literature, such as age, gender, severity, and comorbidity with additional neurodevelopmental disorders (DuPaul et al. 2018; Murray et al. 2014; Zablotsky et al. 2018). More research is needed on the possible relevance of machine learning techniques to identify predictors of the receipt of psychosocial treatment for ADHD.

Present Study

Leveraging machine learning techniques, we explored a range of factors hypothesized to be associated with receiving psychosocial treatment for ADHD in a large, national sample of youth ages 3 to 17. Based on previous empirical evidence and theory (DuPaul et al. 2018; Eiraldi et al. 2006; Murray et al. 2014; Zablotsky et al. 2018; Zendarski et al. 2018), we anticipated that several characteristics possibly associated with *problem recognition* would emerge as correlates of receiving psychosocial treatment: symptom severity, mental health comorbidities. We hypothesized that characteristics possibly associated with *the decision to seek treatment* would emerge as predictors of receiving psychosocial treatment as well: gender (male more likely to receive treatment). We also hypothesized that several factors relevant to *service selection* (e.g., insurance coverage) and *service use* (e.g., receiving special education services) would emerge as related to receiving psychosocial treatment (DuPaul et al. 2018; Murray et al. 2014; Zablotsky et al. 2018; Zendarski et al. 2018).

Method

Data Source and Sample

Data for this study were collected from 2016 to 2017 from households that participated in the National Survey of Children's Health (NSCH; Child and Adolescent Health Measurement Initiative; CAHMI 2019). A data use agreement was completed between the study authors and the Data Resource Center and the Child and Adolescent Health Measurement Initiative (CAHMI). The NSCH was a nationally representative survey on the health and well-being of youth aged 0 to 17 years living in the United States (CAHMI 2017). The NSCH was administered via mail and the internet to randomly selected addresses from civilian, non-institutionalized households (CAHMI 2017). First, instructions were mailed to complete the survey online, and after at least one reminder, those who had not yet accessed the survey were mailed a paper questionnaire. The current study was determined to be exempt from institutional review board review, not categorized as human subjects research because the data were completely de-identified at the time of study.

Participants

Participants were 6630 youth drawn from the 2016–2017 National Survey of Children's Health (NSCH; Child and Adolescent Health Measurement Initiative 2019) with a parent-reported diagnosis of ADHD. With respect to race/ethnicity, our sample was 10.3% Hispanic of any race, 70.3% non-Hispanic white, 7.0 non-Hispanic Black, and 9.5% other race or ethnicity. The age of youth included in our analyses ranged from 3–17 years ($M = 12.4$, $SD = 3.4$), and our sample was 68.5% Male.

Measures

Inclusion Criteria

Children under the age of 3 were excluded from our analyses. To identify children with a diagnosis of ADHD, parents responded to the following question "Has a doctor or other health care provider ever told you that your child had ADHD or attention-deficit disorder (ADD) (even if he or she does not have the condition now)?" Further, parents were also asked "if yes, does this child currently have the condition [ADHD]?" (Child and Adolescent Health Measurement Initiative; CAHMI 2019). In our analyses, we elected to included youth with both past and current ADHD, given that youth with past ADHD were nearly 2 years older than those with a current diagnosis, and nearly one in five teens with ADHD fails to meet the diagnostic criteria for ADHD despite suffering from clinically significant impairment (Sibley et al. 2012). Further, we ran our machine learning analyses with and without these participants: the majority of predictors of psychosocial treatment for ADHD identified by machine learning models remained unchanged, and the predictive accuracy of our models did not significantly differ with or without the inclusion of these participants.

Predictor Variables

The NSCH 2016–2017 survey included 475 items, including factors relating to Eiraldi and colleagues' (2006) help-seeking behavior model (of variables hypothesized to be related to utilization of ADHD treatment). NSCH 2016–2017 variables could be described as relating to problem recognition (e.g., ADHD severity; the extent to which concerns affect the child's ability to do things; comorbid conditions), to the decision to seek help (e.g., demographic characteristics such as age, gender, race/ethnicity; family structure), service selection (e.g., health insurance coverage), and service use (e.g., coordination of care; receipt of other treatments such as medication; Eiraldi et al. 2006). In particular, NSCH 2016–2017 materials describe the variables included in the dataset as pertaining to several content areas: physical and oral health, emotional and mental health, health insurance coverage, health care access and quality, community and school activities, family health and activities, neighborhood safety and support (CAHMI 2018). To view a copy of the survey questionnaire, please go to https://childhealthdata.org/learn/NSCH/resources/survey-instruments.

Outcome Variable

Our primary outcome variable, receipt of psychosocial treatment for ADHD, was evaluated via parent response to the question "At any time DURING THE PAST 12 MONTHS, did this child receive behavioral treatment for ADD or ADHD, such as training or an intervention that you or this child received to help with his or her behavior?" (CAHMI 2019). Youth with missing data (< 1%) on this item were excluded from our analyses.

Data Analytic Plan

Descriptives

Descriptive statistics were completed in SPSS v26 (IBM SPSS 2019).

Machine Learning Analyses

We used the NSCH 2016 ($N=4617$) to train, and the NSCH 2017 ($N=2013$) three machine learning models: classification and regression tree analysis (CART; Breiman 2017); an ensemble model (a random decision forest; Breiman 2017); and a deep, multi-layer neural network (Vincent et al. 2008). Models were run in *Node.js* with the BigML library (BigML 2018). CART is a non-parametric decision tree learning technique that uses recursive partitioning (Breiman 2017). A random decision forest is an ensemble learning method that uses bootstrap-aggregating, "bagging," of decision trees to help address concerns with "overfitting" training data found in most single decision trees (when findings reflect the structure of the training dataset too closely and generate less accurate predictions in other samples; Breiman 2017). We also implemented a deep, multi-layer neural network, which is an algorithm, modeled after human brain, designed to recognize patterns with multiple layers between the input and output layers (Vincent et al. 2008). We selected three non-parametric, machine learning models ranging in various analytical characteristics which can best be described as a trade-off between flexibility and interpretability; CART is the most interpretable, yet the least flexible of the three models (Breiman 2017), and an ensemble model (Friedman 2001) and a neural network (Vincent et al. 2008) have more flexibility, yet less interpretability. Of note, we only included one example of each category of machine learning model in an effort to be concise—one single decision tree (CART), one ensemble model of trees (a random decision forest), and one deep neural net. We also included a logistic regression model for the purposes of comparing machine learning techniques to previously established traditional statistical models. We intend for our analyses to exemplify the promise of the application of machine learning techniques to investigate mental health issues, and we clarify that the models included are *not* exhaustive. For example, we note that we omitted the comparison of additional suitable machine learning models such as chi-square automatic interaction detector (a single decision tree; Kass 1980), or an ensemble model of gradient boosted trees (Friedman 2001). For a thorough review of the application of machine learning techniques to mental health-related data, we refer the reader to Kapczinski et al. (2019).

Of note, we assumed an atheoretical analytical approach, including a total of 668 variables; we aimed to include as many variables in the dataset as possible (of 770 total variables in the NSCH 2016–2017 dataset available for download; CAHMI 2019). We only eliminated variables from analysis due to redundancy in the dataset and changes between NSCH 2016 and 2017 surveys. There were 475 items in the survey, yet 770 variables included in the dataset available for download; for example, certain items were combined to form indicators such as using height and weight to calculate Body Mass Index (which were all retained as predictors in our analyses). An example of using items to calculate entirely redundant variable includes splitting a resilience indicator into two age groups: ages 6 months to 5 years as well as ages 6 to 17 years (we retained the original variable as a predictor in our analyses only). These indicators were likely broken into groups to facilitate calculations for various stakeholders.

Model Evaluation, Interpretation and Computational Efficiency

Receiver operator characteristic (ROC) curves were completed in SPSS v26 (IBM SPSS 2019) to inspect the true positive rate (sensitivity) and false positive rate (1-specificity) of each machine learning model's performance, with the ROC curve figure displayed for our test dataset only (NSCH 2017, $N=2013$). We also note that we elected to keep our data from 2016 and 2017 separate (rather than pooled) in our analytical approach for the purposes of temporal validation (Konig et al. 2007). We evaluated the area under the ROC curve (AUC) of the train dataset (NSCH 2016) with k-fold ($k=10$) cross validation run in in *Node.js* with the BigML library (BigML 2018). The true positive rate (sensitivity) was calculated by dividing the number of true positives by the sum of the number of true positives and the number of false negatives. The false positive rate (1-specificity) was calculated by: 1 minus the number of false positives divided by the sum of the number of true negatives and the number of false positives. True positive instances were defined as the case in which a child with ADHD receives psychosocial treatment for ADHD. Additionally, variable importance figures were computed for all three machine learning models (i.e., mean decrease in impurity, Breiman 2017). More specifically, variable or feature importance figures are an estimate of the ranked indication of the relative significance of input variables to a given machine learning model (Auret and Aldrich 2011). Lastly, we reported the amount of time it took to run each of our models (on a macbook pro laptop computer with standard processor/memory configurations) on the train dataset (NSCH 2016) with k-fold ($k=10$) cross validation to characterize the efficiency of each model.

Results

Percentage of Children with ADHD Receiving Psychosocial Treatment

Of the 6630 total youth in the 2016–2017 NSCH sample with a parent-reported diagnosis of ADHD, 2757 (41.6%) received psychosocial treatment.

Machine Learning Analyses

Receiver operating characteristic curve analysis revealed the following model performance, trained on the NSCH 2016 ($n = 4617$), and tested on the NSCH 2017 data ($n = 2013$): logistic regression (area under the curve; AUC = .69; 95% CI .66–0.71); classification and regression tree analysis (AUC = .68; 95% CI .66–.70); an ensemble model (AUC = .72; 95% CI .70–.75); and a deep, multi-layer neural network (AUC = .71; 95% CI .69–.74). Please see Fig. 1 for a graph of the ROC curve (displaying model performance in the NSCH 2017 test dataset; $n = 2013$). That is, the deep neural network model achieved the highest AUC, though confidence intervals overlap between all three models. To provide information about the fit of our machine learning models to training data, we report that the AUC for 100% our training data only (NSCH 2016; $n = 4617$) with k-fold cross validation (k = 10): (logistic regression; AUC = .58; CART; AUC = .75; ensemble, AUC = .79; deep neural network, AUC = .78). Based on variable importance estimates (Breiman 2017), machine learning models in unanimously identified parent-reported ADHD severity as the factor that best distinguishes between children who receive psychosocial treatment for ADHD and those who do not. Overall, 32.2% of youth with parent-reported mild ADHD received psychosocial treatment for ADHD, vs. 55.4% of youth with parent-reported moderate/severe ADHD, $\chi^2(1, 6022) = 323.9$, $p < .001$, odds ratio = 2.62 (95% CI 2.4, 2.9). Child characteristics such as parent report of the extent to which a child's health problems affected their ability to do things and the number of health conditions a child has differentiated between youth with ADHD who received psychosocial treatment vs, those who did not. Further, factors such as mental/behavioral health insurance coverage and other healthcare service factors, such as receiving medication treatment, satisfaction with healthcare provider communication emerged as predictors of the receipt of psychosocial treatment. We note that variable importance figures were computed for all three models (i.e., mean decrease in impurity, Breiman, 2017), and the top 10 importance estimates for each model (including overlap) are displayed in Table 2. Lastly, we report that the amount of time it took to run each of our models was as follows for the train (NSCH 2016) dataset with k-fold cross validation: logistic regression (67 min); CART (3 min); ensemble (33 min); and deep neural network (44 min). Processing times were faster without running k-fold cross validation (e.g., CART took 17 s).

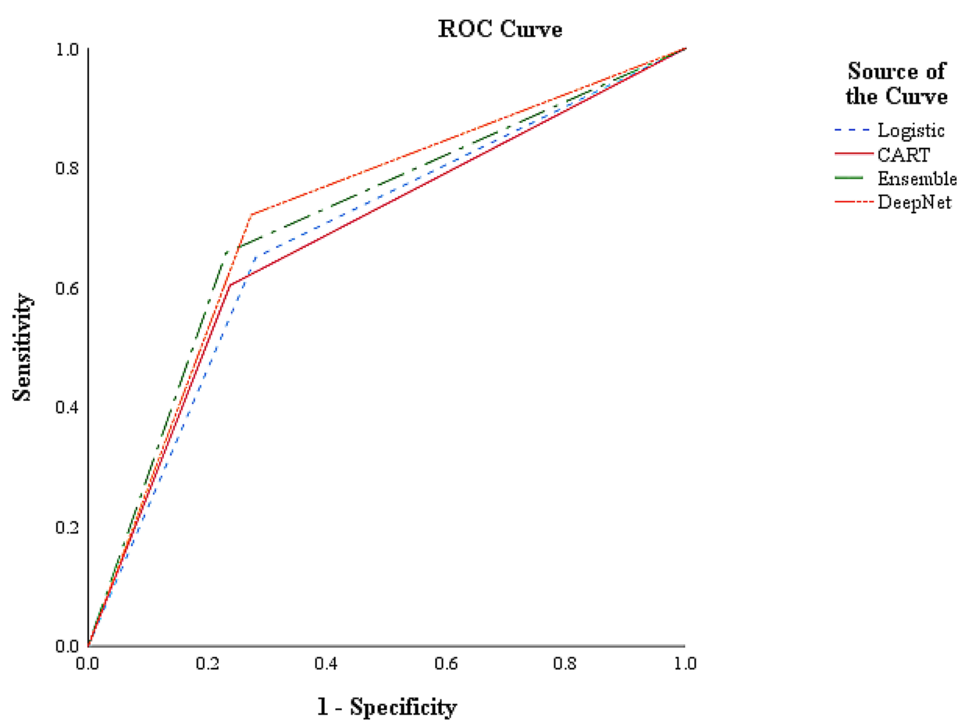

Fig. 1 Receiver Operating Characteristic (ROC) Curve: Machine Learning Models Predicting Receipt of Behavioral Treatment for ADHD. *Logistic* logistic regression, *CART* classification and regression trees, *Ensemble* ensemble decision forest model, *DeepNet* deep, multi-layer neural network

Table 2 Variable/feature importance estimates

Variable	CART	Ensemble	DeepNet
ADHD severity (mild vs. moderate/severe)	51.2	27.6	20.1
Health insurance covers mental/behavioral needs	10.8	9.2	9.6
Receives medication treatment for ADHD	0	18.3	2.9
To what extent do this child's health conditions/problems affect their ability to do things?	2.1	1.6	4.7
Receives behavioral treatment for autism	1.1	0	2.9
Maternal age at birth	0.8	0.6	0
Birthweight in ounces	1.1	0	0
Autism severity (mild, moderate, severe)	0.8	0	0
How often have you felt your child was hard to care for?	0.8	0.6	0
How often get as much help as you wanted with arranging/coordinating this child's health care?	0.7	0	0
Doctors/health care providers worked to create a written plan to meet specific health goals?	0.7	0	0
Age in years of Adult #2 living in the house	0	0.7	0
Age in years of Adult #1 living in the house	0	0.7	0
Parent felt child is much harder to care for than most children during the past month	0	0.6	0
Child age in years	0	0.6	0
Number of health conditions (1 or more vs. 2 or more conditions)	0	0	4.7
Age in years when first given a special education plan	0	0	2.1
Parent report of satisfaction with communication among doctors	0	0	1.6
Child's race	0	0	1.0
Year that Adult #2 living in the house moved to the United States	0	0	1.0
Birth order of selected child in the household	0	0	0.8

CART classification and regression trees, *Ensemble* ensemble decision forest model, *DeepNet* deep, multi-layer neural network. Variable importance figures were computed (i.e., mean decrease in impurity, Breiman 2017) such that all factors included in the model carry an importance weight that sums to 100. The top 10 importance estimates for each model (including overlap) are shown here

Discussion

The current study identified predictors of receiving psychosocial treatment for ADHD in a nationally recruited epidemiological sample of youth ages 3 to 17. Consistent with previous research (DuPaul et al. 2018; Murray et al. 2014), our machine learning models also identified ADHD severity as a predictor of receiving psychosocial treatment for ADHD. Further, mental/behavioral health insurance coverage emerged as a salient predictor of the receipt of psychosocial treatment in our analyses. Several variables identified by previous studies as predictors of receiving services for ADHD, such as gender and lower family income did not emerge in our models as strong predictors of receiving psychosocial treatment (DuPaul et al. 2018). Our study adds to the recent literature on machine learning predictive models that were trained on a relatively large training dataset including individuals with ADHD and tested on data from another, separate cohort (e.g., Caye et al. 2019). Receiver operating characteristic curve analysis revealed performance for three machine learning models ranging from an area under the curve of .68 to .75, as well as comparison to a logistic regression model with an AUC of .69. We discuss each of these findings in more detail below.

Clinical Implications

Less than half of youth in our nationally drawn sample (from 2016 to 2017) with a parent-reported diagnosis of ADHD received psychosocial treatment. In order to examine the current research-to-practice gap, we investigated the correlates of receiving psychosocial treatment for ADHD. Our findings of factors associated with the receipt of ADHD buttress Eiraldi et al. (2006) model of help-seeking behavior for ADHD treatment, encompassing four categories or stages, from problem recognition, to the decision to seek help, service selection, and service utilization patterns. For example, ADHD severity, consistent with previous research (Murray et al. 2014), may affect *problem recognition*. Factors such as insurance coverage (identified in previous literature; DuPaul et al. 2018), could affect *service selection*, as families without access to care covered by insurance could be priced out of behavioral services. Our analyses identified components of s*ervice utilization patterns* as a predictor of receiving psychosocial treatment not yet examined in the previous empirical literature: for example, elements of coordinated care such as satisfaction with communication between healthcare providers predicted receiving psychosocial treatment. In sum, our study added to the theoretical and empirical literature on predictors of receiving psychosocial treatment for ADHD by including strengths such as new methodological techniques (e.g., machine learning), as well as a dataset from a nationally drawn study.

Future Directions for Research and Implications for Policy

The deep neural network achieved the highest area under the curve, followed by the ensemble model (random decision forest), and the classification and regression tree (CART). Notably, the logistic regression model (for comparison to traditional statistical modeling), performed comparably to all three machine learning models. The fastest model to run (CART), was also the least accurate, and the inverse was true for the slowest machine learning model to run (deep neural network), which was the most accurate. However, though the highest AUC is sometimes considered an indicator of the "best" performance, the confidence interval overlapped between all three of our models, meaning the difference in prediction performance were not statistically significantly different between models. Given these findings, there are several important considerations for future research on children's mental health, which could include leveraging a combination of traditional statistical modeling and machine learning approaches, research using "big data" datasets (e.g., MarketScan; Adamson et al. 2008), as well as even machine learning models that are deployed in real time systems.

An exciting direction for future research includes the hope that traditional statistical methods and machine learning could interact in a synergistic way. Although Breiman (2001) described two cultures of statistical modeling (traditional vs. algorithmic/machine learning), perhaps future researchers could become proficient in both spheres. Despite current enthusiasm for machine learning approaches, traditional statistical methods led to many past scientific breakthroughs; traditional models identified important relationships between variables in medicine such as cigarettes and cancer (e.g., Hammond and Horn 1954). Within the field mental health, foundational theory on human behavior (e.g., Social Learning Theory; Bandura and McClelland 1977) relies on the backbone of citations of empirical research derived from traditional statistical methods. In certain situations, traditional statistical models are superior to machine learning models, because they are sometimes more interpretable and leverage the findings from the previous literature to avoid spending computational resources (e.g., the approach illustrated by Youngstrom et al. 2018). Many of the best-performing predictive models currently leverage expert curation of predictor variables, as well as experienced team members to develop and tune predictive models. We optimized our machine learning models to balance the tradeoff between resources spent to develop the models (e.g., time spent to write high-quality computer code) vs. overall performance (i.e., accuracy). Perhaps the optimization of our machine learning models could be helpful as an example of a feasible approach in the context of limited resources (which we note is not a replacement for expert curation), or even the possibility in the future to develop models deployed in real-time systems.

Machine learning could be helpful to apply research findings to improve the outcomes of at-risk youth in real time (e.g., Wisconsin's state-wide high school dropout forecasting system powered by machine learning; Knowles 2015). In the private sector, there have been recent efforts to leverage machine learning to provide real-time, evidence-based feedback to ameliorate mental health problems (Cummins et al. 2019; Fitzpatrick et al. 2017). A machine learning model to analyze therapist speech was developed to ultimately provide prompts to therapists regarding components of cognitive behavioral treatment such as setting homework goals for patients (Cummins et al. 2019). As other fields and sectors leverage machine learning to solve problems, how can stakeholders with an interest in public mental health keep pace? We argue that adding machine learning components to existing digital resources might be a good place to start. For example, the Child and Adolescent Health Measurement Initiative hosts a Data Resource Center webpage with an interactive query available for the NSCH data (available at www.childhealthdata.org/browse/survey). Policy-makers can use the interactive query to calculate descriptive statistics regarding the scope of the problem (e.g., to apply for federal block grants). Perhaps adding a machine learning feature to the interactive query could allow stakeholders to evaluate factors associated with special health needs in their respective catchment areas. That is, private companies leverage machine learning to develop recommender systems (e.g., video-steaming platforms such as Netflix; Gomez-Uribe and Hunt 2015); if a policy-maker using the interactive query were interested in the receipt of psychosocial treatment for ADHD, maybe a "recommender" system pointing him/her to associated variables could be useful to narrow down which components of Eiraldi et al. (2006) ADHD help-seeking behavior model are particularly relevant to address.

Lastly, we highlight that a strength of our study is the implementation of our machine learning models (and logistic regression) in Node.JS., a Javascript runtime environment. There are several advantages that we highlight to our implementation of data analyses in Node.JS (as Javascript has characteristics of an object-oriented language), relative to computer programming languages with a different paradigm, such as functional programming. Object-oriented programming languages typically are better-suited for creating projects that "scale," because they typically have code with less ad-hoc code reuse (an indicator of poor code writing quality) and that takes less time to test (Harrison et al. 1996). That is, object-oriented programming languages could be described as better-suited big data tasks, because of the systematic nature of the development/writing of the code. Systematically written code may encourage collaboration between computer code writers on large tasks written by more than one person. Lastly, Javascript is well-suited to big data tasks because it completes certain benchmark tasks faster than other languages such as R (Bezanson et al. 2012). Node.JS completes tasks quickly due to a feature that allows concurrent processing of more than one task to run at once, called non-blocking input/out. Therefore, Node.JS could be characterized as both lightweight and efficient. Recently, there were investigations of two distinct, large datasets regarding psychosocial treatment for youth with ADHD, each with over 800,000 individuals (OIG 2019; Waxmonsky et al. 2019). Historically, youth mental health research employing much smaller datasets have not necessarily required the use of advanced computer programming languages—though the advances Node.JS (object-oriented, handles multiple concurrent tasks) may be particularly helpful in the future given the growing availability of big data in youth mental health research.

Limitations

One major limitation to our study is the lack of information on what parent-reported psychosocial treatment for ADHD consisted of; one study found, for example, that less than a fourth interventions for ADHD were evidence-based (Murray et al. 2014). Following epidemiological methods, we used parent-reported ADHD status, though a gold-standard, multi-trait, multi-informant assessment would have been ideal (Anastopoulos 2001). Another drawback of our outcome variable (a single, parent-reported item), is that this variable shares that same source/method as critical predictor variables, such as parent-reported symptom severity, which will lead to a phenomenon called shared method variance (i.e., an inflation of the correlation between the criterion and predictors; Campbell and Fiske 1959; Podsakoff et al. 2012). Additionally, our analyses presented here include only comparisons between 3 machine learning models for parsimony; models compared were not exhaustive and future studies should include other analytic techniques such as gradient boosted trees (Friedman, 2001). The description of our variable importance estimates was also not exhaustive; to our knowledge, there is not yet a standardized format for presenting variable importance figures (Breiman 2017). Further, we ideally would have employed several machine learning techniques better-suited to high dimensionality such as regularization (e.g., ridge, LASSO, elastic net). Relatedly, our model lacked an exhaustive strategy to narrow down redundancy among predictor variables, which could lead to statistical artifacts in our findings. Lastly, our test dataset was collected by the same investigators, in the same country as our training dataset. Although the varying time periods (NSCH 2016 vs. NSCH 2017) allowed us to validate our models according to current standards for multivariable predictive models (Moons et al. 2015), it would be helpful to evaluate model performance in a different sample, such as international youth (e.g., Caye et al. 2019).

Summary and Conclusion

In summary, the current study identified factors associated with receiving psychosocial treatment for ADHD in a nationally recruited sample. We found that machine learning models unanimously identified parent-reported ADHD severity as the factor that best distinguishes between children who receive psychosocial treatment for ADHD and those who do not. The area under the receiver operating characteristic curve of our models ranged from .68 (classification and regression tree), to .71 (an ensemble model), to .72 (a deep, multi-layer neural network), and we also included comparison to a logistic regression model with an AUC of .69. Further, mental/behavioral health insurance coverage emerged as a salient factor associated with the receipt of psychosocial treatment. Machine learning models were helpful to investigate predictors of receiving psychosocial treatment for ADHD, confirming previous findings, as well as identifying factors not yet examined in previous empirical research.

Acknowledgements This work was completed while Dr. Morrow was at FIU. She is currently a post-doctoral fellow at Nova Southeastern University.

Compliance with Ethical Standards

Conflict of interest The authors do not have any conflicts of interest to disclose.

Ethical Approval The following study used publicly available data from the National Survey of Children's Health (CAHMI, 2019). All procedures performed in the NSCH study involving human participants were in accordance with the ethical standards of the IRB HHS regulations (45CFR 46), these procedures are reviewed by the NHS Research Ethics Review Board (ERB) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Research Involving Human Rights The current study was determined to be exempt from institutional review board review at Florida International University, not categorized as human subjects research because the data were completely de-identified at the time of study.

References

Adamson, D. M., Chang, S., & Hansen, L. G. (2008). *Health research data for the real world: The MarketScan databases* (p. b28). New York: Thompson Healthcare.

Aguiar, E., Lakkaraju, H., Bhanpuri, N., Miller, D., Yuhas, B., & Addison, K. L. (2015). Who, when, and why: A machine learning approach to prioritizing students at risk of not graduating high school on time. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 93–102).https://doi.org/10.1145/2723576.2723619

American Academy of Pediatrics: Subcommittee on ADHD, steering committee of Quality Improvement and Management (2011). ADHD: Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents. *Pediatrics*, *128*(5), 1007–1022. https://doi.org/10.1542/peds.2011-2654.

Altszuler, A. R., Page, T. F., Gnagy, E. M., Coxe, S., Arrieta, A., Molina, B. S. G., et al. (2015). Financial dependence of young adults with childhood ADHD. *Journal of Abnormal Child Psychology*. https://doi.org/10.1007/s10802-015-0093-9.

Anastopoulos, A. D. (2001). *Assessing attention-deficit/hyperactivity disorder* (2001st ed.). New York: Springer.

Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). *Predicting student dropout in higher education*. Retrieved from https://arxiv.org/abs/1606.06364v4.

Auret, L., & Aldrich, C. (2011). Empirical comparison of tree ensemble variable importance measures. *Chemometrics and Intelligent Laboratory Systems, 105*(2), 157–170. https://doi.org/10.1016/j.chemolab.2010.12.004.

Bandura, A., & McClelland, D. C. (1977). *Social learning theory*. Retrieved from https://www.esludwig.com/uploads/2/6/1/0/26105457/bandura_sociallearningtheory.pdf.

Bezanson, J., Karpinski, S., Shah, V. B., & Edelman, A. (2012). Julia: A fast dynamic language for technical computing. *ArXiv:1209.5145* [Cs]. Retrieved from https://arxiv.org/abs/1209.5145

BigML. (2018). Retrieved from https://bigml.com. BigML, Inc.: Corvallis, OR.

Bowers, A. J., & Zhou, X. (2019). Receiver operating characteristic (ROC) area under the curve (AUC): A diagnostic measure for evaluating the accuracy of predictors of education outcomes. *Journal of Education for Students Placed at Risk, 24*(1), 20–46. https://doi.org/10.1080/10824669.2018.1523734.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science, 16*(3), 199–231. https://doi.org/10.1214/ss/1009213726.

Breiman, L. (2017). Classification and regression trees. *Routledge*. https://doi.org/10.1201/9781315139470.

Burke, T. A., Ammerman, B. A., & Jacobucci, R. (2019). The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. *Journal of Affective Disorders, 245*, 869–884. https://doi.org/10.1016/j.jad.2018.11.073.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Caye, A., Agnew-Blais, J., Arseneault, L., Gonçalves, H., Kieling, C., Langley, K., et al. (2019). A risk calculator to predict adult attention-deficit/hyperactivity disorder: Generation and external validation in three birth cohorts and one clinical sample. *Epidemiology and Psychiatric Sciences, 29*, 1–9.

Child and Adolescent Health Measurement Initiative. (2017). 2016 National Survey of Children's Health (2017), Sampling and Survey Administration. Data Resource Center, supported by Cooperative Agreement 1-U59-MC06980-01 from the U.S. Department of Health and Human Services, Health Resources and Services Administration (HRSA), Maternal and Child Health Bureau (MCHB). Retrieved from www.childhealthdata.org. Revised 04/26/17.

Child and Adolescent Health Measurement Initiative. (2018). *Child and Family Health Measures Content Map, 2017 National Survey of Children's Health*. Data Resource Center for Child and Adolescent Health supported by Cooperative Agreement U59MC27866 from the U.S. Department of Health and Human Services, Health Resources and Services Administration's Maternal and Child Health Bureau (HRSA MCHB). Retrieved from www.childhealthdata.org, Revised 9/26/2018

Child and Adolescent Health Measurement Initiative (CAHMI). (2019). *2016–2017 National Survey of Children's Health, SPSS Indicator Data Set*. Data Resource Center for Child and Adolescent Health supported by Cooperative Agreement from the U.S. Department of Health and Human Services, Health Resources and Services Administration (HRSA), Maternal and Child Health Bureau (MCHB). Retrieved 15 April, 2019, from www.childhealthdata.org.

Cummins, R., Ewbank, M. P., Martin, A., Tablan, V., Catarino, A., & Blackwell, A. D. (2019). TIM: A tool for gaining insights into psychotherapy. *The World Wide Web Conference*. https://doi.org/10.1145/3308558.3314128.

Danielson, M. L., Bitsko, R. H., Ghandour, R. M., Holbrook, J. R., Kogan, M. D., & Blumberg, S. J. (2018). Prevalence of parent-reported ADHD diagnosis and associated treatment among US children and adolescents, 2016. *Journal of Clinical Child and Adolescent Psychology, 47*(2), 199–212. https://doi.org/10.1080/15374416.2017.1417860.

DuPaul, G. J., Chronis-Tuscano, A., Danielson, M. L., & Visser, S. N. (2018). Predictors of receipt of school services in a national sample of youth with ADHD. *Journal of Attention Disorders*. https://doi.org/10.1177/1087054718816169.

Eiraldi, R. B., Mazzuca, L. B., Clarke, A. T., & Power, T. J. (2006). Service utilization among ethnic minority children with ADHD: A model of help-seeking behavior. *Administration and Policy in Mental Health and Mental Health Services Research, 33*(5), 607–622. https://doi.org/10.1007/s10488-006-0063-1.

Evans, S. W., Owens, J. S., Wymbs, B. T., & Ray, A. R. (2018). Evidence-based psychosocial treatments for children and adolescents with Attention Deficit/Hyperactivity Disorder. *Journal of Clinical Child & Adolescent Psychology, 47*(2), 157–198. https://doi.org/10.1080/15374416.2017.1390757.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health, 4*(2), e19. https://doi.org/10.2196/mental.7785.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451.

Gomez-Uribe, C. A., & Hunt, N. (2015). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems, 6*(4), 13–19. https://doi.org/10.1145/2843948.

Hammond, E. C., & Horn, D. (1954). The relationship between human smoking habits and death rates: A follow-up study of 187,766 men. *Journal of the American Medical Association, 155*(15), 1316–1328. https://doi.org/10.1001/jama.1954.03690330020006.

Harrison, R., Samaraweera, L. G., Dobie, M. R., & Lewis, P. H. (1996). Comparing programming paradigms: An evaluation of functional and object-oriented programs. *Software Engineering Journal, 11*(4), 247–254.

IBM Corp. (2019). *IBM SPSS statistics for windows, version 26.0*. Armonk, NY: IBM Corp.

Huber, B., Davis III, R. F., Cotter, A., Junkin, E., Yard, M., Shieber, S., Brestan-Knight, E., & Gajos, K. Z. (2019). SpecialTime: Automatically detecting dialogue acts from speech to support parent-child interaction therapy. *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 139–148.

Jensen, P. S., Hinshaw, S. P., Swanson, J. M., Greenhill, L. L., Conners, C. K., Arnold, L. E., et al. (2001). Findings from the NIMH multimodal treatment study of ADHD (MTA): Implications and applications for primary care providers. *Journal of Developmental & Behavioral Pediatrics, 22*(1), 60.

Kapczinski, F., Passos, I. C., & Mwangi, B. (2019). *Personalized psychiatry: Big data analytics in mental health*. New York: Springer.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 29*(2), 119–127. https://doi.org/10.2307/2986296.

Kessler, R. C., Bernecker, S. L., Bossarte, R. M., Luedtke, A. R., McCarthy, J. F., Nock, M. K., et al. (2019). The role of big data analytics in predicting suicide. In I. C. Passos, B. Mwangi, & F. Kapczinski (Eds.), *Personalized psychiatry: Big data analytics in mental health* (pp. 77–98). New York: Springer. https://doi.org/10.1007/978-3-030-03553-2_5.

Kim, J.-W., Sharma, V., & Ryan, N. D. (2015). Predicting methylphenidate response in ADHD using machine learning approaches. *The International Journal of Neuropsychopharmacology, 18*(11), pyv052. https://doi.org/10.1093/ijnp/pyv052.

Knowles, J. E. (2015). Of needles and haystacks: Building an accurate statewide dropout early warning system in Wisconsin. *Journal of Educational Data Mining, 7*(3), 18–67.

Konig, I. R., Malley, J. D., Weimar, C., Diener, H. C., Ziegler, A., & German Stroke Study, C. (2007). Practical experiences on the necessity of external validation. *Statistics in Medicine, 26*, 5499–5511.

Kuriyan, A. B., Pelham, W. E., Molina, B. S. G., Waschbusch, D. A., Gnagy, E. M., Sibley, M. H., et al. (2013). Young adult educational and vocational outcomes of children diagnosed with ADHD. *Journal of Abnormal Child Psychology, 41*(1), 27–41. https://doi.org/10.1007/s10802-012-9658-z.

Lenhard, F., Sauer, S., Andersson, E., Månsson, K. N., Mataix-Cols, D., Rück, C., et al. (2018). Prediction of outcome in internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: A machine learning approach. *International Journal of Methods in Psychiatric Research, 27*(1), e1576. https://doi.org/10.1002/mpr.1576.

Merrill, B. M., Molina, B. S. G., Coxe, S., Gnagy, E. M., Altszuler, A. R., Macphee, F. L., et al. (2019). Functional outcomes of young adults with childhood ADHD: A latent profile analysis. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division, 53*, 1–14. https://doi.org/10.1080/15374416.2018.1547968.

Moons, K. G., Altman, D. G., Reitsma, J. B., Ioannidis, J. P., Macaskill, P., Steyerberg, E. W., et al. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine, 162*(1), W1–W73.

Murray, D. W., Molina, B. S. G., Glew, K., Houck, P., Greiner, A., Fong, D., et al. (2014). Prevalence and characteristics of school services for high school students with Attention-Deficit/Hyperactivity Disorder. *School Mental Health, 6*(4), 264–278. https://doi.org/10.1007/s12310-014-9128-6.

Ng, X., Bridges, J. F. P., Ross, M. M., Frosch, E., Reeves, G., Cunningham, C. E., et al. (2017). A latent class analysis to identify variation in caregivers' preferences for their child's Attention-Deficit/Hyperactivity Disorder treatment: Do stated preferences match current treatment? *The Patient, 10*(2), 251–262. https://doi.org/10.1007/s40271-016-0202-z.

Office of Inspector General (OIG). (2019). Many Medicaid-enrolled children who were treated for ADHD did not receive recommended follow-up care. U.S. Department of Health and Human Services: Washington, DC. Retrieved from oig.hhs.gov/oei/reports/oei-07-17-00170.asp.

Page, T. F., Pelham, W. E., Fabiano, G. A., Greiner, A. R., Gnagy, E. M., Hart, K. C., et al. (2016). Comparative cost analysis of sequential, adaptive, behavioral, pharmacological, and combined treatments for childhood ADHD. *Journal of Clinical Child and Adolescent Psychology, Division, 45*(4), 416–427. https://doi.org/10.1080/15374416.2015.1055859.

Pelham, W. E. P., Jr., Fabiano, G. A., Waxmonsky, J. G., Greiner, A. R., Gnagy, E. M., III, Pelham, W. E., et al. (2016). Treatment sequencing for childhood ADHD: A multiple-randomization study of adaptive medication and behavioral interventions. *Journal of Clinical Child & Adolescent Psychology, 45*(4), 396–415. https://doi.org/10.1080/15374416.2015.1105138.

Pérez-Rosas, V., Mihalcea, R., Resnicow, K., Singh, S., & An, L. (2016). Building a motivational interviewing dataset. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology* (pp. 42–51). https://doi.org/10.18653/v1/W16-0305

Pérez-Rosas, V., Mihalcea, R., Resnicow, K., Singh, S., Ann, L., Goggin, K. J., et al. (2017). Predicting counselor behaviors in motivational interviewing encounters. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pp. 1128–1137. Retrieved from https://aclweb.org/anthology/E17-1106

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development, 3*(3), 210–229.

Sansone, D. (2019). Beyond early warning indicators: High school dropout and machine learning. *Oxford Bulletin of Economics and Statistics, 81*(2), 456–485. https://doi.org/10.1111/obes.12277.

Sara, N.-B., Halland, R., Igel, C., & Alstrup, S. (2015). High-school dropout prediction using machine learning: A Danish large-scale study. *European Symposium on Artificial Neural Networks*.

Schatz, N., Fabiano, G., Cunningham, C., dosReis, S., Waschbusch, D., Jerome, S., et al. (2015). Systematic review of patients' and parents' preferences for ADHD treatment options and processes of care. *The Patient: Patient-Centered Outcomes Research, 8*(6), 483–497.

Sibley, M. H., Pelham, W. E., Molina, B. S. G., Gnagy, E. M., Waschbusch, D. A., Garefino, A. C., et al. (2012). Diagnosing ADHD in adolescence. *Journal of Consulting and Clinical Psychology, 80*(1), 139–150. https://doi.org/10.1037/a0026577.

Subcommittee on Attention-Deficit, Hyperactivity Disorder, S. C. on Q. I. & M. (2011). ADHD: Clinical practice guideline for the diagnosis, evaluation, and treatment of Attention-Deficit/Hyperactivity Disorder in children and adolescents. *Pediatrics*. https://doi.org/10.1542/peds.2011-2654.

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. *In Proceedings of the 25th International Conference on Machine Learning* (pp. 1096–1103). New York, NY: ACM. https://doi.org/10.1145/1390156.1390294

Waschbusch, D. A., Cunningham, C. E., Pelham, W. E., Rimas, H. L., Greiner, A. R., Gnagy, E. M., et al. (2011). A discrete choice conjoint experiment to evaluate parent preferences for treatment of young, medication naive children with ADHD. *Journal of Clinical Child and Adolescent Psychology The Official Journal for the*

Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division, 40(4), 546–561. https://doi.org/10.1080/15374416.2011.581617.

Waxmonsky, J. G., Baweja, R., Liu, G., Waschbusch, D. A., Fogel, B., Leslie, D., et al. (2019). A commercial insurance claims analysis of correlates of behavioral therapy use among Children With ADHD. *Psychiatric Services*. https://doi.org/10.1176/appi.ps.201800473.

Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology, 39*(2), 204–221. https://doi.org/10.1093/jpepsy/jst062.

Youngstrom, E. A., Halverson, T. F., Youngstrom, J. K., Lindhiem, O., & Findling, R. L. (2018). Evidence-based assessment From simple clinical judgments to statistical learning: Evaluating a range of options using pediatric bipolar disorder as a diagnostic challenge. *Clinical Psychological Science, 6*(2), 243–265. https://doi.org/10.1177/2167702617741845.

Zablotsky, B., Bramlett, M. D., Visser, S. N., Danielson, M. L., & Blumberg, S. J. (2018). Latent class analysis of ADHD neurodevelopmental and mental health comorbidities. *Journal of Developmental and Behavioral Pediatrics, 39*(1), 10–19. https://doi.org/10.1097/DBP.0000000000000508.

Zendarski, N., Sciberras, E., Mensah, F., & Hiscock, H. (2018). Factors associated with educational support in young adolescents with ADHD. *Journal of Attention Disorders*. https://doi.org/10.1177/1087054718804351.

Zhao, X., Page, T. F., Altszuler, A. R., Pelham, W. E., Kipp, H., Gnagy, E. M., et al. (2019). Family burden of raising a child with ADHD. *Journal of Abnormal Child Psychology*. https://doi.org/10.1007/s10802-019-00518-5.