A precision health approach to medication management in neurodevelopmental conditions: a model development and validation study using four international cohorts

Marlee M. Vandewouw[1,2,3] PhD (ORCID: 0000-0002-4505-0306); Kamran Niroomand[1], MEng (ORCID: 0000-0003-1937-8372); Harshit Bokadia[1], MSc (ORCID: 0000-0002-5929-079X); Sophia Lenz[1], MS.; Jesiqua Rapley[1], MA; Alfredo Arias[4], BA; Jennifer Crosbie[5,6], PhD (ORCID: 0000-0002-8710-3322); Elisabetta Trinari[7], MD (ORCID: 0000-0001-5859-652X); Elizabeth Kelley[8], PhD (ORCID: 0000-0001-7742-6542); Robert Nicolson[9], MD (ORCID: 0000-0001-7086-7038); Russell J. Schachar[5,6], MD (ORCID: 0000-0002-2015-4395); Paul D. Arnold[10], MD (ORCID: 0000-0003-2496-4624); Alana Iaboni[1], PhD. Jason P. Lerch[11,12,13], PhD (ORCID: 0000-0001-6164-2881); Melanie Penner[1], MD (ORCID: 0000-0002-8376-9768); Danielle Baribeau[1, 5,6,14,15], MD (ORCID: 0000-0003-4461-5851); Evdokia Anagnostou[1,14,16], MD (ORCID: 0000-0002-3455-9887); Azadeh Kushki[1,17], PhD (ORCID: 0000-0003-1510-8818)

1. Autism Research Center, Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, Canada
2. Center for Precision Psychiatry, Massachusetts General Hospital, Boston, United States of America
3. Department of Psychiatry, Harvard Medical School, Boston, United States of America
4. Province of Ontario Neurodevelopmental (POND) Network, Toronto, Canada
5. Department of Psychiatry, The Hospital for Sick Children, Toronto, Canada
6. Department of Psychiatry, University of Toronto, Toronto, Canada
7. Faculty of Health Sciences, McMaster University, Hamilton, Canada
8. Departments of Psychology and Psychiatry, Queen's University, Kingston, Canada
9. Department of Psychiatry, University of Western Ontario, London, Canada
10. Department of Psychiatry, Cumming School of Medicine, University of Calgary, Calgary, Canada
11. Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom
12. Department of Medical Biophysics, University of Toronto, Toronto, Canada
13. Program in Neurosciences & Mental Health, The Hospital for Sick Children, Toronto, Canada
14. Institute of Medical Sciences, University of Toronto, Toronto, Canada
15. Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada
16. Department of Paediatrics, University of Toronto, Toronto, Canada
17. Institute of Biomedical Engineering, The University of Toronto, Toronto, Canada

Corresponding author:
Dr. Azadeh Kushki
Bloorview Research Institute
Holland Bloorview Kids Rehabilitation Hospital
150 Kilgour Road Toronto, Ontario M4G 1R8
akushki@hollandbloorview.ca
Tel: 416-425-6220, extension 3563

Abstract

Background: Psychotropic medications are commonly used for children with neurodevelopmental conditions, but their effectiveness varies, making prescribing challenging and potentially exposing individuals to multiple medication trials. This study developed artificial intelligence (AI) models to predict prescribing patterns of stimulants, anti-depressants, and anti-psychotics.

Methods: Models were built to predict cross-sectional medication usage from the Child Behaviour Checklist. Data from the Province of Ontario Neurodevelopmental (POND) network (*N*=598) trained and tested the models, while data from the Healthy Brain Network (HBN; *N*=1,764) and Adolescent Brain Cognitive Development (ABCD; *N*=2,396) studies were used for external validation. For the electronic medical record (EMR) cohort, data from the Psychopharmacology Program (PPP; *N*=312) at Holland Bloorview Kids Rehabilitation Hospital were used to predict longitudinal success. Stacked ensemble models were built separately for each medication class, and area under the receiver operating characteristic curve (AU-ROC) evaluated performance.

Results: The research cohorts demonstrated feasibility, with internal testing (POND) achieving an AU-ROC (median [IQR]) of 0.75 [0.73,0.80] for stimulants, 0.83 [0.78,0.87] for anti-depressants, and 0.79 [0.72,0.86] for anti-psychotics. Performance in external testing sets (HBN and ABCD) confirmed generalizability. In the EMR cohort (PPP), AU-ROCs were 0.84 [0.81,0.88] for stimulants, 0.82 [0.77,0.87] for anti-depressants, and 0.87 [0.83,0.91] for anti-psychotics.

Limitations: The study's findings should be interpreted with caution due to several factors, including the use of medication choice as a proxy for treatment response and the omission of key sociodemographic variables, which may introduce bias. The model was trained solely on North American data, limiting global generalizability, and predictions were restricted to broad medication classes rather than within-class choices.

Conclusions: Findings demonstrate that AI can accurately learn expert prescribing patterns and predict medication prescribing decisions, supporting the potential of data-

driven tools to guide personalized medication management for neurodevelopmental conditions and reducing the trial-and-error burden in clinical practice.

Keywords: artificial intelligence, medication management, neurodevelopmental conditions, autism

Background

Neurodevelopmental conditions, such as autism spectrum disorder (autism), intellectual disability (ID), and attention-deficit/hyperactivity disorder (ADHD), are highly heterogeneous and overlapping in their biology and phenotypic presentation[1,2], which includes hundreds of associated genes[3], and significant variability in neurobiological correlates[2]. Despite decades of research, no markers exist to indicate which biological pathway may be implicated for an individual child to guide individualized care[4,5], and current approaches to care provision remain reliant on broad diagnostic labels. This lack of precision is particularly challenging for pharmacological interventions in neurodevelopmental conditions. These interventions are highly prevalent with greater than 50% of children with neurodevelopmental conditions prescribed psychotropic medications to address behavioural and/or mental health concerns (e.g., anxiety, hyperactivity, aggressive or self-injurious behavior), and high rates of polypharmacy during childhood (~30%)[6].

Despite the prevalent use of psychotropic medications in neurodevelopmental conditions, there is tremendous variability in individual-level responses to psychotropic treatment for co-occurring mental health and behavioural conditions[7], reflecting the heterogeneity of biological pathways impacted, diverse clinical presentations, as well as sociodemographic complexities that influence the acceptability and uptake of medications. This variability poses a significant challenge in determining the appropriate medication for an individual child, especially in cases where co-occurring conditions (e.g., autism, ADHD, anxiety) can interact and increase the complexity of presentation and response. In these cases, prescribing practices in the children and adolescents remain almost entirely off-label[6,8–12] and based on trial-and-error approaches that can expose children to multiple medication trials with the potential for severe and life-long adverse effects for some classes of medication[13]. This highlights the critical need for precision health tools that can provide medication recommendations based on an individual’s unique presentation.

To fill this gap, we propose an artificial intelligence (AI) model that uses an individual’s characteristics to predict the medication choice among the commonly prescribed classes in neurodevelopmental conditions: stimulants, anti-depressants, and

anti-psychotics. We first establish the feasibility of this approach by building models to predict medication usage using a research cohort of children and youth with neurodevelopmental conditions, using two international research cohorts for external validation. Then, to examine how this approach generalizes to real-world clinical settings, we leveraged a patient cohort to predict continuation of a medication class after 3 months from electronic medical record (EMR) data of children and youth with neurodevelopmental conditions who visited a specialized psychopharmacology clinic.

Methods

Study design and cohorts

For the research cohorts, training and internal testing was conducted using data from 598 children and youth with neurodevelopmental conditions extracted from the Province of Ontario Neurodevelopmental (POND) network (Ontario, Canada). External testing was performed using two international cohorts: 1,764 children and youth with neurodevelopmental conditions from the Healthy Brain Network (HBN[14]; New York, United States), and 2,396 children and youth with neurodevelopmental conditions from the Adolescent Brain Cognitive Development study (ABCD[15]; United States).

For clinical validation, longitudinal EMRs from the Psychopharmacology Program (PPP) at Holland Bloorview Kids Rehabilitation Hospital (Toronto, Ontario, Canada) were used. PPP sees children with neurodevelopmental conditions who have tried at least one psychoactive medication without adequate success. Data were extracted from electronic medical records of 312 children who visited the PPP over a three-year period (2019 – 2022) and returned for a follow-up approximately 3 months later.

Participant demographics are summarized in

Table 1. Further details on the cohorts, including eligibility and consent procedures, and a comparison of the research cohorts can be found in the Supplementary Material. This study was conducted in accordance with the Declaration of Helsinki, and was approved by the Research Ethics Board at Holland Bloorview Kids Rehabilitation Hospital (#1854; Clinical trial number: not applicable).

Table 1. Cohort demographics and medication outcomes.

		Research cohorts			EMR cohort
		POND	**HBN**	**ABCD**	**PPP**
N		598	1,764	2,396	312
Age range (years)		5 – 19	5 – 17	8 – 11	4 – 17
Mean age [SD] (years)		11.4 [3.2]	10.1 [2.7]	9.9 [0.6]	10.4 [2.9]
Sex [%]	Male	424 [71]	1,213 [69]	1,523 [64]	262 [84]
	Female	174 [29]	551 [31]	873 [36]	50 [16]
Autism [%]		338 [57]	325 [18]	0 [0]	292 [94]
ADHD [%]		193 [32]	1,381 [78]	1,751 [73]	181 [58]
OCD [%]		67 [11]	90 [5]	1,048 [44]	16 [5]
Intellectual disability [%]	Yes	84 [14]	86 [5]	149 [6]	105 [34]
	No	514 [86]	1,678 [95]	2,247 [94]	135 [43]
	Unknown	0 [0]	0 [0]	0 [0]	72 [23]
Race/ ethnicity [%]	White	327 [55]	871 [49]	1,257 [52]	NA
	Minoritized	156 [26]	866 [49]	1,137 [47]	NA
	Not collected	115 [19]	27 [2]	2 [0]	312 [100]
Annual household income [%]	Low	86 [14]	307 [17]	770 [32]	NA
	Medium	98 [16]	345 [20]	608 [25]	NA
	High	170 [28]	786 [45]	798 [33]	NA
	Not collected	244 [41]	326 [18]	220 [9]	312 [100]
Outcome [%]	Stimulant	171 [29]	177 [10]	551 [23]	114 [37]
	Anti-depressant	96 [16]	72 [4]	142 [6]	82 [26]
	Anti-psychotic	52 [9]	26 [1]	43 [2]	123 [39]

EMR: electronic medical record; PPP: Psychopharmacology Program; POND: Province of Ontario Neurodevelopmental network; HBN: Healthy Brain Network; ABCD: Adolescent Brain Cognitive Development study; SD: standard deviation; ADHD: attention-deficit/hyperactivity disorder; OCD: obsessive-compulsive disorder. For POND, primary diagnosis is reported; for ABCD, HBN, and PPP, refer to supplemental.

Predictors and outcome

For the PPP cohort, the set of predictors were co-created with the clinicians at the PPP program to reflect features relevant to the medication classes considered, and the availability of variables in the electronic medical records at the first visit. These predictors were broadly grouped into five categories: demographics, clinical, family-related, child-related, and care-related, the last of which included prior use of psychotropic medication classes at the first visit. The predicted outcome was the medication decision at the follow-up visit, focusing on three primary classes: stimulants, anti-depressants, and anti-psychotics (

Table 1). Non-stimulants were considered in a supplementary analysis, given variability in the clinical classification of this medication group. A full description of the predictors and outcome are presented in the Supplementary Material.

For the research cohorts, demographics (age and sex), neurodevelopmental diagnoses (autism, ADHD, and OCD), IQ as a continuous score, and data from the Child Behaviour Checklist (CBCL)[16] were used as predictors. The predicted outcome was the medication class (stimulants, anti-depressants, anti-psychotics) the child was on at the time of study (

Table 1). A full description of the predictors and outcome are presented in the Supplementary Material, and a between-cohort comparison of the demographics and CBCL scores are reported in Supplemental Tables 1 and 2.

Model

AI models were built to predict the outcome from the set of predictors, separately for each medication class (**Error! Reference source not found.**). For the research cohorts, the models were trained and internally tested using the POND dataset, while HBN and ABCD were withheld for external testing. By validating with two external datasets, we assessed the generalizability and robustness of the models, ensuring results are not specific to a single dataset and can be replicated across diverse populations. For the EMR cohort (PPP), only internal testing was performed.

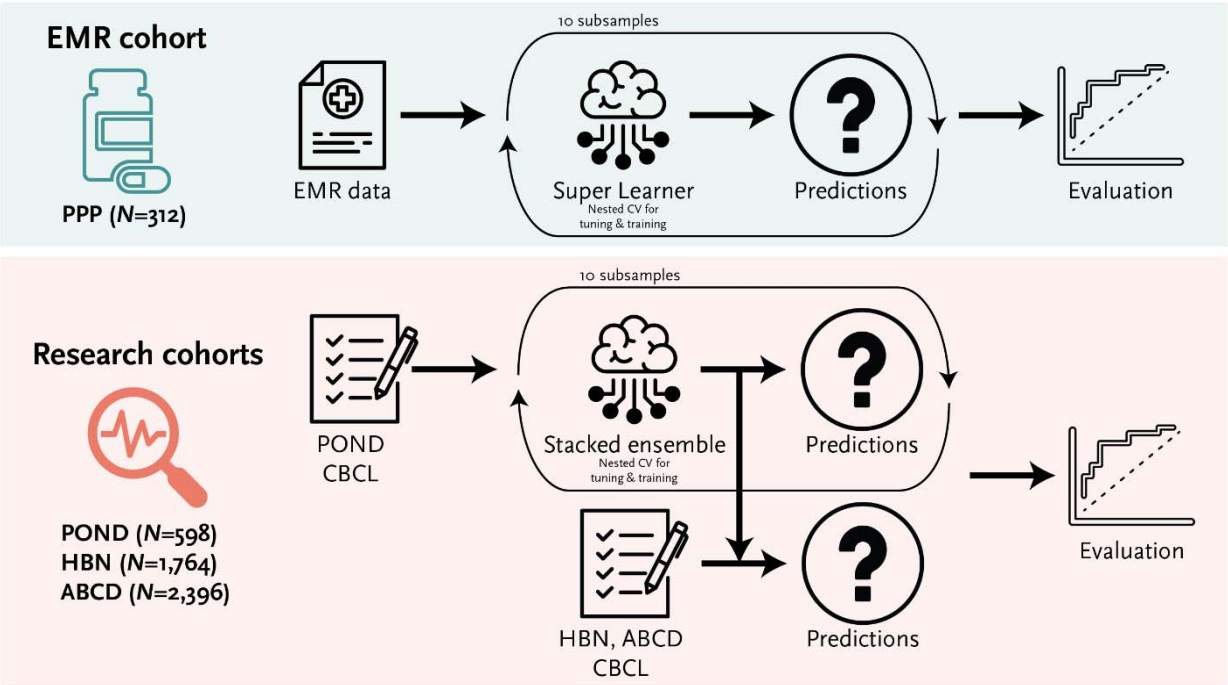

Figure 1. Study design.

To address class imbalance, we applied random under-sampling of the majority class to achieve approximately balanced 1:1 class ratios. This algorithm-agnostic approach reduces majority-class dominance and strengthens the minority-class learning signal without relying on synthetic data or class weighting. To limit dependence on any

single balanced draw, we repeated the under-sampling for 10 independent iterations, noting that findings were stable across 5, 10, and 20 iterations.

For the AI models, we implemented a Super Learner approach[17], combining CatBoost gradient boosting[18] and deep neural networks as base learners, with logistic regression meta-learner trained on out-of-fold predictions to generate final predictions. We used nested cross-validation with inner 3-fold loops for hyperparameter optimization using Optuna[19] and outer 5-fold loops where base learners generated out-of-fold predictions for the logistic regression meta-learner. Further details are provided in the Supplementary material.

Performance evaluation

Model performance was assessed using the area under the receiver operating characteristic curve (AU-ROC), reporting medians and interquartile ranges (IQR) across folds and under-sampling iterations. Additional metrics (area under the precision-recall curve (AU-PR), sensitivity, and specificity) are provided in the Supplementary Material.

To interpret which features were most important for predicting longitudinal medication choice, Shapley Additive exPlanations (SHAP) values[20] were computed for the EMR (PPP) models. SHAP values were computed for each of the base models and combined by taking their weighted average using the coefficients from the logistic regression meta-learner.

The demographic parity and equalized odds ratios[21,22] were used to evaluate whether the model predictions were independent of sensitive group status for sex, intellectual disability (ID), race/ethnicity, and annual household income. The metric was computed at the default probability threshold (0.50) across folds, under-sampling iterations, and datasets for the research cohorts; the 80% rule was used to determine any biases[23].

Results

Model performance

The research cohorts demonstrated the feasibility of AI predicting treatment status cross-sectionally for each medication class (**Error! Reference source not found.**), with internal testing (POND) demonstrating good-to-excellent AU-ROC that was significantly higher than chance for stimulants (median [IQR]: 0.75 [0.73, 0.80]),

anti-depressants (0.83 [0.78, 0.87]), and anti-psychotics (0.79 [0.72, 0.86]). The performance of the external testing sets (HBN and ABCD) indicated good generalizability, with the AU-ROCs slightly reduced but still in the good-to-excellent range. Additional metrics (area under the precision-recall curve, sensitivity, specificity) are reported in Supplemental Table 3. For the EMR cohort (PPP; **Error! Reference source not found.**A), the AU-ROC across the cross-validation folds was excellent for stimulants (0.84 [0.81, 0.88]), anti-depressants (0.82 [0.77, 0.87]), and anti-psychotics (0.87 [0.83, 0.91]). Predictive performance for non-stimulants is reported in Supplemental Table 4.

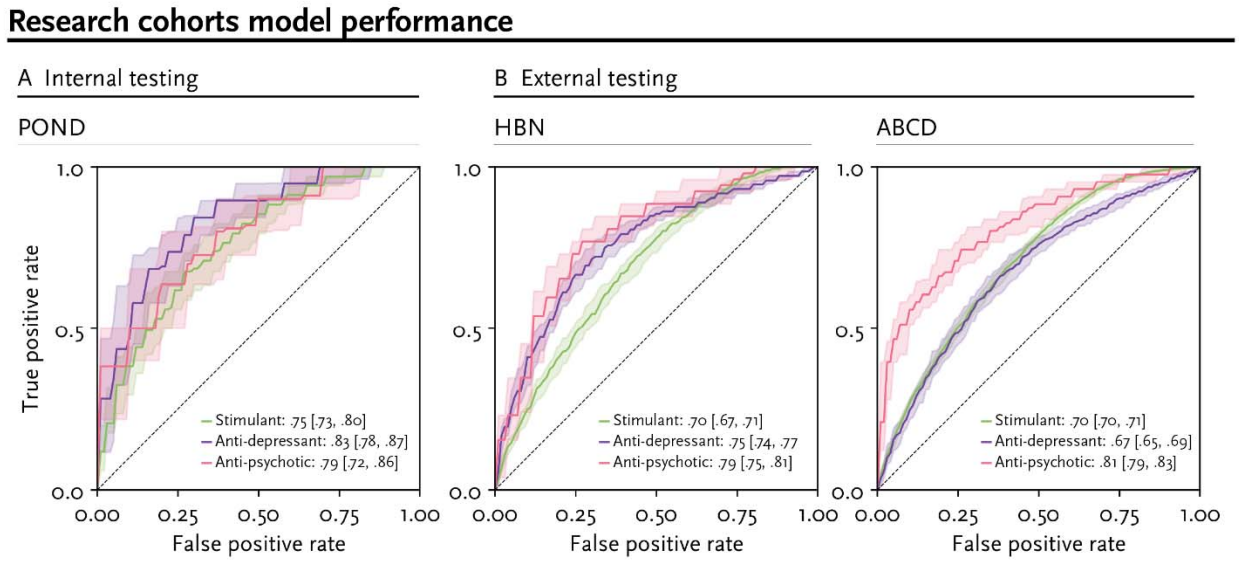

Figure 2. ROC curves for the research cohorts, for the internal testing (A; POND) and external testing (B; HBN and ABCD) sets for each of the medication classes (green: stimulants, purple: anti-depressants, pink: anti-psychotics); AU-ROCs are also presented.

POND: Province of Ontario Neurodevelopmental network; HBN: Healthy Brain Network; ABCD: Adolescent Brain Cognitive Development study; ROC: receiver operating characteristic curve; AU-ROC: area under the ROC

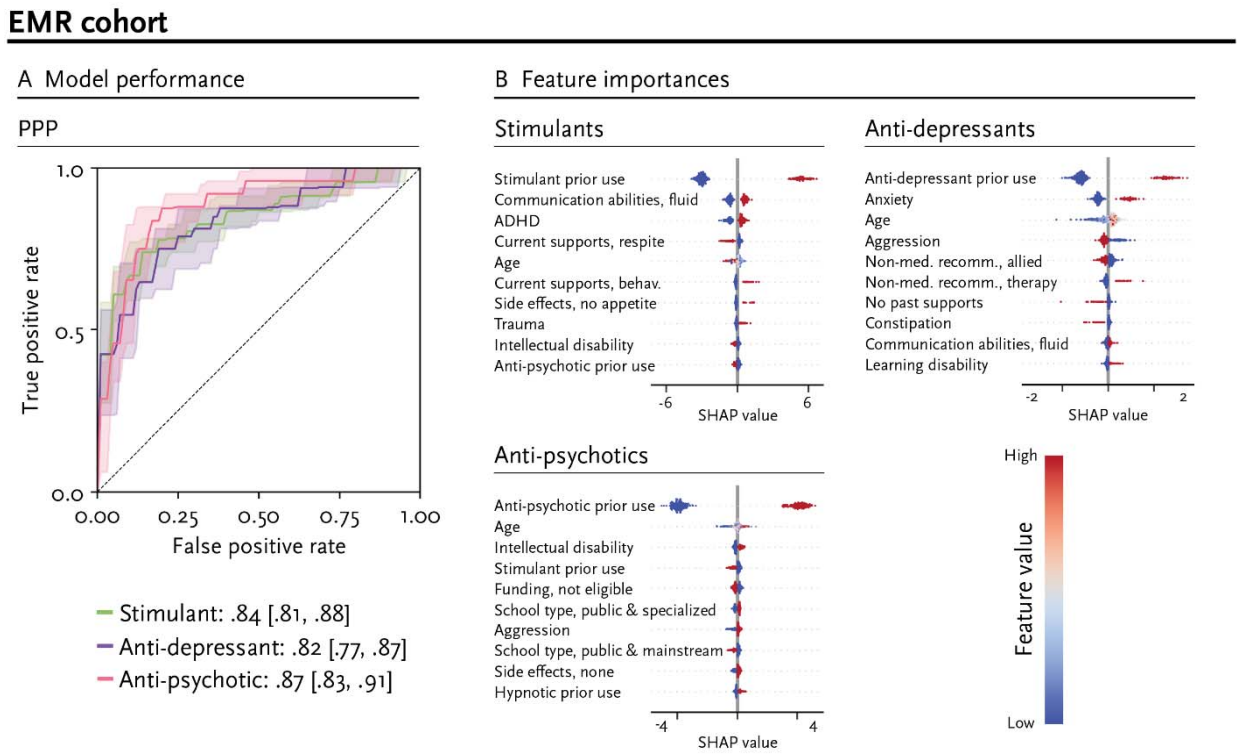

Figure 3. (A) ROC curves for the EMR cohort (PPP) for each of the medication classes (green: stimulants, purple: anti-depressants, pink: anti-psychotics); AU-ROCs are also presented. (B) Feature importance for the EMR cohort (PPP) assessed using SHAP values. The SHAP values are plotted for each participant for the 10 most important features for each medication class, sorted by importance (mean absolute SHAP value) from top to bottom. Colours represent the feature value (red for high values, blue for low values), indicating whether higher or lower values contribute positively or negatively to prediction scores.

EMR: electronic medical record; PPP: Psychopharmacology Program; ROC: receiver operating characteristic curve; AU-ROC: area under the ROC; SHAP: Shapley Additive exPlanations; ADHD: attention-deficit/hyperactivity disorder

Feature importance

SHAP values were computed for the EMR cohort (PPP), to identify which features were most important for predicting medication choice (**Error! Reference source not found.**B). Prior use of a medication class was the most important feature across all three classes. In particular, prior use of a medication within the same class, increased the likelihood of a medication choice in that class. Age was also a salient feature: younger, middle, and older participants were more likely to be recommended stimulants, anti-psychotics, and anti-depressants, respectively. Other important features included ADHD and communication abilities for stimulants, anxiety and aggression for anti-depressants, and intellectual disability and prior prescription of stimulants for anti-

psychotics. In a sensitivity analysis, re-fitting of the EMR model after removing the prior medication-use features reduced AU-ROCs as expected, but all models remained significantly above chance (Supplemental Table 5), indicating that the remaining predictors contribute independent signal.

Bias

The demographic parity and equalized odds ratios were computed to evaluate the bias of the models (Table 2), where values below 0.80 indicate bias according to the four-fifths rule[24]. To support interpretation of the ratios, per-group metrics are provided in Supplemental Tables 6 and 7 for the research and EMR cohorts, respectively. For the research cohorts, parity was mixed. Individuals who were male, and who did not have an intellectual disability (ID) were more likely to be recommended a stimulant (female, male: 39%, 56%; ID, no-ID: 37%, 52%). Individuals who were female, did not have an ID, who were white, and who were of high income class were more likely to be recommended an anti-depressant (female, male: 40%, 27%; ID, no-ID: 23%, 32%; white, minority: 38%, 22%; low, middle, high income: 31%, 25%, 35%). On the other hand, those with an ID, and those with a lower annual household income were more likely to be recommended anti-psychotics (ID, no-ID: 54%, 31%; low, middle, high income: 45%, 30%, 22%). These observations extended to the EMR cohort, with the exception of sex, with males being more likely to be recommended stimulants (male, female: 45%, 36%) and females being more likely to be recommended anti-depressants (male, female: 44%, 55%). This was in line with trends in the data.

Table 2. Demographic parity ratio (DPR) and equalized odds ratio (EOR) for sensitive groups.

		Research cohorts		EMR cohort	
		DPR	**EOR**	**DPR**	**EOR**
Sex	Stimulant	0.70	0.70	0.79	0.49
	Anti-depressant	0.69	0.68	0.79	0.72
	Anti-psychotic	0.81	0.83	0.91	0.81
Intellectual disability	Stimulant	0.72	0.63	0.45	0.29
	Anti-depressant	0.71	0.65	0.62	0.41
	Anti-psychotic	0.57	0.74	0.62	0.59
Race/ethnicity	Stimulant	0.95	0.94	–	–
	Anti-depressant	0.57	0.71	–	–
	Anti-psychotic	0.83	0.90	–	–
Annual household income (low vs. middle)	Stimulant	0.96	0.85	–	–
	Anti-depressant	0.82	0.89	–	–
	Anti-psychotic	0.67	0.60	–	–
Annual household income (low vs. high)	Stimulant	0.94	0.85	–	–
	Anti-depressant	0.87	0.86	–	–
	Anti-psychotic	0.50	0.56	–	–
Annual household income (middle vs. high)	Stimulant	0.98	0.91	–	–
	Anti-depressant	0.72	0.77	–	–
	Anti-psychotic	0.75	0.66	–	–

EMR: electronic medical record

To determine whether observed disparities reflect or amplify existing prescribing patterns, we compared the model-predicted DPR to baseline DPR computed from actual prescribing labels (Supplemental Table 8). In the research cohorts, the model moved DPR above the four-fifths threshold for several attributes where baseline prescribing was disparate, including anti-psychotic predictions by sex (baseline DPR=0.56, model DPR=0.81), stimulant and anti-psychotic predictions by race/ethnicity (baseline DPR = 0.66 and 0.29, model DPR = 0.95 and 0.83), and stimulant predictions by income (baseline DPR = 0.76, model DPR = 0.94 for the most disparate pair). Conversely, the model introduced disparities for intellectual disability where baseline prescribing showed near parity for stimulants (baseline DPR = 0.97, model DPR = 0.72) and anti-depressants (baseline DPR = 0.89, model DPR = 0.71). In the EMR cohort, no attribute meaningfully crossed the four-fifths threshold in either direction.

Discussion

Medication management for children with neurodevelopmental conditions remains a significant challenge, particularly in community settings where subspecialty expertise may not be available. To address this gap, this study aimed to develop and evaluate an AI algorithm to predict medication choices in neurodevelopmental conditions. Ultimately, this type of tool could be used as a decision-aid to support medication management where specialized expertise is not available.

We showed that AI can learn expert clinician medication decisions with high accuracy. We also demonstrated that the results generalize across independent, international datasets, which is notable given the drastically different characteristics of these cohorts, in terms of demographics, diagnosis, phenotype, and region-specific access to service and supports (see **Error! Reference source not found.** and Supplemental Table 2). While the accuracy in these research cohorts is lower compared to the EMR model, this is likely due to the cross-sectional nature of the prediction in these cohorts. For example, use of stimulant medications to target ADHD behaviours confounds the measures of ADHD traits when considered cross-sectionally, in turn impacting predictive performance.

Overall, these results support the feasibility of using an AI-based decision aid tool. This can be particularly advantageous for community practitioners, where subspecialty expertise required for medication management is often not available. Ultimately, an AI-based decision aid can bring precision to current prescribing practices in the community which may expose children to multiple medication trials that may have severe and life-long adverse effects[13]. Prospective trials are needed to further validate this type of algorithms in clinical settings.

Explainability and interpretability are critical features of responsible AI systems, contributing to enhancing trust and transparency with both clinicians and patients[25]. To promote explainability, we characterized the relative contributions of clinical and demographic features to the AI algorithm's medication recommendations. For the EMR

cohort, these contributions were closely aligned with indications for each medication class (e.g., stimulants for ADHD, SSRIs for anxiety). Interestingly, the top contributors for all classes were prior use of a medication in that class. This finding reflects the fact that medication choices often remain within the same class of medication but entail changes in the medication type and dosing within that class (e.g., amphetamines, methylphenidate within the stimulant class). Future work will focus on adding this type of resolution to the AI predictions. Although prior use was the strongest contributor, a sensitivity analysis re-fitting the model without any of the prior use features showed that performance decreased but remained significantly above chance across all three classes. This is expected given that PPP, by design, serves children who have already tried one or more psychoactive medications – making prior exposure a core input to every clinical decision at this program – but also indicates that the remaining demographic, clinical, family-related, child-related, and care-related features carry independent predictive signal beyond current medication status. Other contributing features to medication decision reflected sociodemographic and environmental factors that may contribute to clinical decision making beyond primary indications. For example, school environment and funding eligibility were significant contributors to the decision to recommend an anti-psychotic. These findings highlight the complexity of the clinical and sociodemographic factors that contribute to value-laden decisions around medication management for children with neurodevelopmental conditions.

Another aspect of responsible AI is potential algorithmic biases that may disproportionately disadvantage subgroups of individuals. Our analysis revealed several biases including increased recommendation of anti-psychotics to children who had intellectual disability and were from lower income households. On the other hand, anti-depressants recommendations were increased for children who did not have an intellectual disability, and who were white. It is important to note that these biases are not related to any specific feature of the AI algorithm, but mirror previously reported biases in prescription patterns which exist across the system[26]. For example, decreased access to early diagnosis and behavioural interventions for children from lower income household may contribute to a higher likelihood of crisis situations that

necessitate prescription of anti-psychotic medications. Similarly, it is well established that depression can be challenging to detect in individuals with an intellectual disability[27].

Comparison of model-predicted to baseline prescribing disparities confirmed that, while the model attenuated disparities related to sex, race/ethnicity, and income in the research cohorts, it introduced new disparities for children with intellectual disability for stimulant and anti-depressant predictions. This may partly reflect differential presentation on behavioural measures such as the CBCL. However, no attribute showed amplified disparity in the EMR cohort, which most closely reflects the intended clinical population for real-world implementation. This is particularly relevant for community settings – the intended deployment context for this type of tool – which often serve children from more marginalized backgrounds. Because the model is trained on historical prescribing data, there is a risk that it could reproduce existing inequities; however, our baseline comparison demonstrated that the model did not amplify most existing disparities, and in several cases attenuated them. Computational bias mitigation strategies, such as fairness constraints during model training and group-specific calibration, can be applied to further reduce predicted disparities, though these technical approaches alone cannot resolve the broader system-level inequities that drive disparate prescribing patterns.

Limitations

The results of this study should be interpreted considering several limitations. Firstly, for the longitudinal prediction, we defined outcome as the medication choice at the follow-up visit. Treatment decisions in neurodevelopmental conditions are a complex, subjective, and value-laden construct. The use of medication choice was co-created with clinicians as an outcome as it reflects the aggregate effects of the multitude of factors contributing to a clinical decision to not stop a medication (e.g., ineffectiveness, side-effects, acceptability or other factors). These children also visited the specialized clinic after at least one psychoactive medication was tried without success, and thus this may introduce biases when the models are used to make a first-line medication

decision. Although gender identity and race/ethnicity were not included as predictors due to incomplete capture in the primary EMR cohort, their omission may contribute to model misspecification bias. These variables remain important contextual factors for understanding medication access and prescribing patterns. Future work should incorporate reliably captured measures and evaluate model performance across gender and racial/ethnic groups, alongside fairness assessments, updating the model as needed to mitigate potential inequities. The models were trained using data exclusively from North America; future work should investigate their translation to other global populations to account for potential differences in healthcare systems, prescribing practices, and sociocultural influences on medication usage. Finally, this study exclusively focused on predicting medication classes; these classes are broad, and future work should examine within-class predictability.

Conclusions

In summary, the present study focused on AI model development and evaluation, demonstrating the feasibility of this approach for medication recommendation for children with neurodevelopmental conditions. To enable effective clinical implementation of this type of AI-tool, future studies are needed to evaluate the algorithm in a prospective design at the point of care as well as acceptability, and barriers and facilitators to uptake.

List of abbreviations

ABCD: Adolescent Brain Cognitive Development; ADHD: attention-deficit/hyperactivity disorder; AI: artificial intelligence; AU-PR: area under the precision-recall curve; AU-ROC: area under the receiver operating characteristic curve; CBCL: Child Behaviour Checklist; DPR: demographic parity ratio; EMR: electronic medical record; EOR: equalized odds ratio; HBN: Healthy Brain Network; ID: intellectual disability; IQR: interquartile range; OCD: obsessive-compulsive disorder; POND: Province of Ontario Neurodevelopmental; PPP: Psychopharmacology Program; ROC: receiver operating characteristic; SD: standard deviation; SHAP: Shapley Additive exPlanations; SSRI: selective serotonin reuptake inhibitor.

Declarations

Ethics approval and consent to participate

This study was approved by the Research Ethics Board at Holland Bloorview Kids Rehabilitation Hospital. As this project involved secondary use of de-identified clinical data, individual participant consent was not required in accordance with institutional and ethical guidelines.

Consent for publication

Not applicable.

Availability of data and materials

Data from four cohorts were used in this study, each governed by their own data sharing restrictions; full details can be found in the Supplementary Material. Code used in this manuscript is available at GitHub (https://github.com/marlvan/med_prediction).

Competing interests

EA is Editor-in-Chief of Molecular Autism. EA has received research funding from Roche and Anavex, consulting fees from Roche, Quadrant Therapeutics, Ono, and Impel Pharmaceuticals, in-kind support from AMO Pharma and CRA-Simons Foundation, royalties from APPI and Springer, an editorial honorarium from Wiley, and has a patent for holly™. DB has received research funding from MapLight Therapeutics. MP has received consulting fee from Addis & Associates/Roche, and MedCounsel. AK has a patent for holly™ with royalties paid from Awake Labs and has received consulting fees from DNAStack and Shaftesbury. The remaining authors declare that they have no competing interests.

Funding

Funding was provided by the Canadian Institutes of Health Research (Operating Grant #527447).

Authors' contributions

Conceptualization: MMV, MP, DB, EA, AK. Methodology: MMV, KN, HB. Software: MMV, KN. Validation: MMV. Formal analysis: MMV. Resources: JC, ET, EK, RN, RJS, PDA, JPL, MP, DB, EA, AK. Data Curation: SL, JR. Writing - Original Draft: MMV, AK. Writing - Review & Editing: MMV, KN, HB, SL, JR, AA, JC, ET, EK, RN, RJS, PDA, AI,

JPL, MP, DB, EA, AK. Visualization: MMV. Supervision: AK. Project administration: AK. Funding acquisition: MMV, MP, DB, EA, AK.

Acknowledgements

Acknowledgements for each of the research cohorts (POND, HBN, and ABCD) used in this study are provided below.

POND: This research was conducted with the support of the Ontario Brain Institute (POND, PIs: Dr. Evdokia Anagnostou and Dr. Jason P. Lerch), an independent non-profit corporation, funded partially by the Ontario government. The opinions, results and conclusions are those of the authors and no endorsement by the Ontario Brain Institute is intended or should be inferred.

HBN: This manuscript was prepared using a limited access dataset obtained from the Child Mind Institute Biobank, the Healthy Brain Network (HBN). This manuscript reflects the views of the authors and does not necessarily reflect the opinions or views of the Child Mind Institute.

ABCD: Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive DevelopmentSM (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this

report came from DOI: 10.15154/z563-zd24. DOIs can be found at https://nda.nih.gov/abcd

References

1. Astle DE, Holmes J, Kievit R, Gathercole SE. Annual Research Review: The transdiagnostic revolution in neurodevelopmental disorders. Journal of Child Psychology and Psychiatry [Internet]. 2021 [cited 2022 Jan 19];n/a. https://doi.org/10.1111/jcpp.13481

2. Vandewouw MM, Brian J, Crosbie J, Schachar RJ, Iaboni A, Georgiades S, et al. Identifying Replicable Subgroups in Neurodevelopmental Conditions Using Resting-State Functional Magnetic Resonance Imaging Data. JAMA Network Open. 2023;6:e232066. https://doi.org/10.1001/jamanetworkopen.2023.2066

3. Trost B, Thiruvahindrapuram B, Chan AJS, Engchuan W, Higginbotham EJ, Howe JL, et al. Genomic architecture of Autism Spectrum Disorder from comprehensive whole-genome sequence annotation [Internet]. Genetic and Genomic Medicine; 2022 May. https://doi.org/10.1101/2022.05.05.22274031

4. Persico AM, Ricciardello A, Lamberti M, Turriziani L, Cucinotta F, Brogna C, et al. The pediatric psychopharmacology of autism spectrum disorder: A systematic review - Part I: The past and the present. Prog Neuropsychopharmacol Biol Psychiatry. 2021;110:110326. https://doi.org/10.1016/j.pnpbp.2021.110326

5. Lord C, Charman T, Havdahl A, Carbone P, Anagnostou E, Boyd B, et al. The Lancet Commission on the future of care and clinical research in autism. Lancet. 2022;399:271–334. https://doi.org/10.1016/S0140-6736(21)01541-5

6. Feroe AG, Uppal N, Gutiérrez-Sacristán A, Mousavi S, Greenspun P, Surati R, et al. Medication Use in the Management of Comorbidities Among Individuals With Autism Spectrum Disorder From a Large Nationwide Insurance Database. JAMA Pediatrics. 2021;175:957–65. https://doi.org/10.1001/jamapediatrics.2021.1329

7. Baribeau D, Vorstman J, Anagnostou E. Novel treatments in autism spectrum disorder. Curr Opin Psychiatry. 2022;35:101–10. https://doi.org/10.1097/YCO.0000000000000775

8. Murray ML, Hsia Y, Glaser K, Simonoff E, Murphy DGM, Asherson PJ, et al. Pharmacological treatments prescribed to people with autism spectrum disorder (ASD) in primary health care. Psychopharmacology (Berl). 2014;231:1011–21. https://doi.org/10.1007/s00213-013-3140-7

9. Lake JK, Weiss JA, Dergal J, Lunsky Y. Child, parent, and service predictors of psychotropic polypharmacy among adolescents and young adults with an autism spectrum disorder. J Child Adolesc Psychopharmacol. 2014;24:486–93. https://doi.org/10.1089/cap.2014.0011

10. Spencer D, Marshall J, Post B, Kulakodlu M, Newschaffer C, Dennen T, et al. Psychotropic medication use and polypharmacy in children with autism spectrum disorders. Pediatrics. 2013;132:833–40. https://doi.org/10.1542/peds.2012-3774

11. Mellahn OJ, Knott R, Tiego J, Kallady K, Williams K, Bellgrove MA, et al. Understanding the Diversity of Pharmacotherapeutic Management of ADHD With Co-occurring Autism: An Australian Cross-Sectional Survey. Front Psychiatry. 2022;13:914668. https://doi.org/10.3389/fpsyt.2022.914668

12. Ritter C, Hewitt K, McMorris CA. Psychotropic Polypharmacy Among Children and Youth with Autism: A Systematic Review. J Child Adolesc Psychopharmacol. 2021;31:244–58. https://doi.org/10.1089/cap.2020.0110

13. Howes OD, Rogdaki M, Findon JL, Wichers RH, Charman T, King BH, et al. Autism spectrum disorder: Consensus guidelines on assessment, treatment and research from the British Association for Psychopharmacology. J Psychopharmacol. 2018;32:3–29. https://doi.org/10.1177/0269881117741766

14. Alexander LM, Escalera J, Ai L, Andreotti C, Febre K, Mangone A, et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci Data. 2017;4:170181. https://doi.org/10.1038/sdata.2017.181

15. Barch DM, Albaugh MD, Avenevoli S, Chang L, Clark DB, Glantz MD, et al. Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. Dev Cogn Neurosci. 2018;32:55–66. https://doi.org/10.1016/j.dcn.2017.10.010

16. Achenbach TM, Rescorla LA. Manual for the ASEBA school-age forms and profiles. Burlington, VT: University of Vermont Research Center for Children, Youth, and Families; 2001.

17. van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol. 2007;6:Article25. https://doi.org/10.2202/1544-6115.1309

18. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc.; 2018. p. 6639–49.

19. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining [Internet]. New York, NY, USA: Association for Computing Machinery; 2019 [cited 2026 Apr 22]. p. 2623–31. https://doi.org/10.1145/3292500.3330701

20. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc.; 2017. p. 4768–77.

21. Bird S, Dudík M, Edgar R, Horn B, Lutz R, Milan V, et al. Fairlearn: A toolkit for assessing and improving fairness in AI [Internet]. Microsoft; 2020 May. Report No.:

MSR-TR-2020-32. https://www.microsoft.com/en-us/research/publication/fairlearn-a-toolkit-for-assessing-and-improving-fairness-in-ai/

22. Agarwal A, Beygelzimer A, Dudik M, Langford J, Wallach H. A Reductions Approach to Fair Classification. In: Dy J, Krause A, editors. Proceedings of the 35th International Conference on Machine Learning [Internet]. PMLR; 2018. p. 60–9. https://proceedings.mlr.press/v80/agarwal18a.html

23. Feldman M, Friedler SA, Moeller J, Scheidegger C, Venkatasubramanian S. Certifying and Removing Disparate Impact. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. New York, NY, USA: Association for Computing Machinery; 2015 [cited 2025 Aug 20]. p. 259–68. https://doi.org/10.1145/2783258.2783311

24. Holzer H, Neumark D. Assessing Affirmative Action. Journal of Economic Literature. 2000;38:483–568. https://doi.org/10.1257/jel.38.3.483

25. Cutillo CM, Sharma KR, Foschini L, Kundu S, Mackintosh M, Mandl KD, et al. Machine intelligence in healthcare-perspectives on trustworthiness, explainability, usability, and transparency. NPJ Digit Med. 2020;3:47. https://doi.org/10.1038/s41746-020-0254-2

26. Antoniou T, McCormack D, Kitchen S, Pajer K, Gardner W, Lunsky Y, et al. Geographic variation and sociodemographic correlates of prescription psychotropic drug use among children and youth in Ontario, Canada: a population-based study. BMC Public Health. 2023;23:85. https://doi.org/10.1186/s12889-022-14677-6

27. Buckley N, Glasson EJ, Chen W, Epstein A, Leonard H, Skoss R, et al. Prevalence estimates of mental health problems in children and adolescents with intellectual disability: A systematic review and meta-analysis. Aust N Z J Psychiatry. 2020;54:970–84. https://doi.org/10.1177/0004867420924101