Looking for Childhood-Onset Schizophrenia: Diagnostic Algorithms for Classifying Children and Adolescents with Psychosis

Deanna Greenstein, PhD,¹ Rachna Kataria, MBBS,¹ Peter Gochman, MA,¹
Abhijit Dasgupta, PhD,² James D. Malley, PhD,³ Judith Rapoport, MD,¹ and Nitin Gogtay, MD¹

Abstract

Objective: Among children < 13 years of age with persistent psychosis and contemporaneous decline in functioning, it is often difficult to determine if the diagnosis of childhood onset schizophrenia (COS) is warranted. Despite decades of experience, we have up to a 44% false positive screening diagnosis rate among patients identified as having probable or possible COS; final diagnoses are made following inpatient hospitalization and medication washout. Because our lengthy medication-free observation is not feasible in clinical practice, we constructed diagnostic classifiers using screening data to assist clinicians practicing in the community or academic centers.

Methods: We used cross-validation, logistic regression, receiver operating characteristic (ROC) analysis, and random forest to determine the best algorithm for classifying COS ($n = 85$) versus histories of psychosis and impaired functioning in children and adolescents who, at screening, were considered likely to have COS, but who did not meet diagnostic criteria for schizophrenia after medication washout and inpatient observation ($n = 53$). We used demographics, clinical history measures, intelligence quotient (IQ) and screening rating scales, and number of typical and atypical antipsychotic medications as our predictors.

Results: Logistic regression models using nine, four, and two predictors performed well with positive predictive values > 90%, overall accuracy > 77%, and areas under the curve (AUCs) > 86%.

Conclusions: COS can be distinguished from alternate disorders with psychosis in children and adolescents; greater levels of positive and negative symptoms and lower levels of depression combine to make COS more likely. We include a worksheet so that clinicians in the community and academic centers can predict the probability that a young patient may be schizophrenic, using only two ratings.

Introduction

PSYCHOTIC SYMPTOMS ARE NOT UNCOMMON IN CHILDHOOD. For example, 13% of a cohort of 11-year-olds reported a psychotic symptom (Poulton et al. 2000). In a cohort of 12-year-old twins, 5.9% reported definite psychotic symptoms, with auditory and visual hallucinations as the most common symptoms (Polanczyk et al. 2010). A third study found a 13.7% 6 month prevalence rate for psychotic symptoms in 12-year-olds (Horwood et al. 2008). As such, it is argued that psychosis exists on a dimension ranging from mild and fleeting to severe, impairing, and persistent (van Os et al. 2009).

In contrast to the frequency of minor psychotic experiences, childhood onset schizophrenia (COS), with psychosis onset before age 13, is a severe, chronic, and rare form of schizophrenia. Given the rarity of COS, there is a high probability that childhood hallucinations and delusions are not necessarily related to COS. At the same time, a portion of children will experience persistent positive symptoms with

additional cognitive and behavioral problems that are not caused by schizophrenia (Bartels-Velthuis et al. 2011). Importantly, some of these children may be moderately or severely impaired such that it may be difficult to determine if they have schizophrenia. We have previously reported on a sample of such individuals enrolled in our longitudinal COS study, who presented with poor functioning and reports of persistent psychotic symptoms but did not ultimately meet diagnostic criteria for schizophrenia (McKenna et al. 1994; Nicolson et al. 2001; Stayer et al. 2005; Gochman et al. 2011).

Over the past two decades, we have conducted > 350 in-person screenings after initially ruling out almost 10 times as many unscreened referrals. As of 2011, of the 217 children and adolescents admitted to our study, 56% ultimately met American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) diagnostic criteria for schizophrenia after an inpatient hospitalization that typically lasted 2–5 months and included a 1–3 week medication washout period. This means that despite extensive

¹Child Psychiatry Branch, National Institute of Mental Health (NIMH)/National Institutes of Health (NIH), Bethesda, Maryland.

²National Institute of Arthritis and Musculoskeletal and Skin Diseases, (NIAMS)/NIH, Bethesda, Maryland.

³Center for Information Technology (CIT)/NIH, Bethesda, Maryland.

Statistical experts: Deanna Greenstein, Abhijit Dasgupta, and James D. Malley.

chart reviews and in-person interviews with children and their parents conducted by experts in the field, up to 44% of the patients we admitted with a provisional diagnosis of schizophrenia ultimately did not meet diagnostic criteria following inpatient hospitalization and medication washout (Gochman et al. 2011). For a clinician who has never encountered a COS patient, it is possible that the risk of making a false positive COS diagnosis could be much higher.

This report is our effort to expand on our previous work (Gochman et al. 2011) with an extended sample, and to create an internally valid, accurate, accessible diagnostic algorithm with measures of sensitivity, specificity, and positive and negative predictive value that would maximize predictive information and could be applied to children and adolescents who fall at the severe end of the psychosis spectrum and present as diagnostic challenges in the community and in academic centers (including our own). This algorithm would allow us to reduce, quantify, and share >20 years of experience, and our unique advantage of having a drug-free inpatient observation period. As such, we applied univariate (one predictor) and multiple predictor methods to determine an algorithm that could most accurately predict whether or not a child or adolescent had schizophrenia. We used screening measures as predictors in order to approximate the experience of clinicians who do not have the opportunity to assess a patient in a medication-free state and/or an inpatient setting. We also included measures associated with risk of schizophrenia such as intelligence quotient (IQ) (Horwood et al. 2008); developmental language; social, motor, and academic problems; and severity of positive and negative symptoms (Asarnow and Benmeir 1988; Hollis 1995; Nicolson and Rapoport 1999).

Methods

Screening assessment and diagnosis

Our screening procedure began with an initial referral and subsequent review of all available medical, academic, and psychological records and a phone screening by a social worker. If there was a documented history of persistent psychosis with onset before 13 years of age accompanied by a decline in functioning, full scale IQ >70, and absence of neurological disease, we invited the potential proband and the proband's family to the National Institutes of Health (NIH) in Bethesda, Maryland for an intensive 2 day in-person screening.

During the in-person screening, staff members conducted unstructured and semistructured (Schedule for Affective Disorders and Schizophrenia for School Age Children) (Kaufman et al. 1997) clinical interviews with potential probands and their legal guardians. Based on information gathered during interviews, staff psychiatrists and social workers completed the Global Assessment of Functioning Scale (GAS) (American Psychiatric Association 1994), the Brief Psychiatric Rating Scale (BPRS) (Lukoff et al. 1986) (intraclass correlation coefficient [ICC]=0.68, 0.70) (Hafkensheid 1993), Scale for the Assessment of Positive Symptoms (SAPS) (Andreasen 1984) (ICC for total score=0.834) (Norman et al. 1996), scale for the Assessment of Negative Symptoms (SANS) (Andreasen 1983) (Chronbach α for total score=0.885) (Andreasen 1982), and the NIMH Global Scale, (Murphy et al. 1982; Sunderland et al. 1988) which is a modification of the Bunney–Hamburg (Bunney and Hamburg 1963) with its own guidelines for rating severity, and which includes separate global depression (ICC=0.67), psychosis (ICC=0.78), mania (ICC=0.64), and anxiety (ICC=0.47) scores rated on a scale of 1–15 (Sunderland et al., 1988). Our raters scored the NIMH Global Scale ratings as 0 in the complete absence of relevant symptoms. We used total scores (sum of all items) from SANS, BPRS, and SAPS, so that higher scores reflected greater symptomatology. GAS and NIMH Global Scale

measures are each a single rating: higher NIMH Global Scale scores reflect greater severity, lower GAS scores reflect worse functioning. Over the course of the study, we calculated ICCs for clinical ratings from contemporaneous raters. ICCs for different groups of child psychiatrists and clinical social workers at five different points of the study ranged from moderate (0.60–0.70) to high (>0.80) with the exception of one ICC between two raters for the NIMH Global Scale Anxiety rating (<0.50). Because of the study's lengthy time span, there is the possibility for differences among groups of raters; however, we ensured that several raters bridged groups of raters to attenuate possible biases related to study year.

Based on information from the 2 day screening, we excluded 38% of individuals (Gochman et al. 2011) because of the sporadic nature of the psychosis (e.g., occurring only under stress) or lack of impairment. For the remaining patients (62% of screened patients), we made a provisional diagnosis of schizophrenia and accepted them to the study. Once accepted, participants were offered an inpatient hospitalization stay and a medication washout, after which their final diagnoses were made. The children and adolescents who did not meet DSM criteria for schizophrenia after inpatient hospitalization and medication washout were classified as having an alternate diagnosis (AD). All diagnoses were made via consensus among staff clinicians after the washout period.

We obtained IQ scores either from an age-appropriate version of the Wechsler Intelligence Scale for Children (WISC) or the Wechsler Abbreviated Scale of Intelligence (WASI) administered while children were inpatients at NIH or from pre-NIH screening post-psychosis onset clinical records. We conducted in-depth chart reviews and interviews for evidence of premorbid impairments in academic, speech/language, motor, and social functioning.

We obtained informed consent from all legal guardians and informed assent from all minor participants at the beginning of the 2 day screening. The study was approved by the institutional review board (IRB).

Participants

For the current study, our sample included 85 COS and 53 AD participants. Requirements for inclusion in the current study included completion of an inpatient washout period, and having complete screening, demographic, IQ, and clinical data. The later inclusion criterion was necessary because predictive models will drop an individual who does not have complete data. Although most participants included in the current study had complete data, occasional items within a screening measure (e.g., SAPS, SANS) were not scored, and in these cases, we used the average from the completed items ($n=4$ COS and $n=2$ AD). Also, several COS participants had complete data except for one rating scale, and in these cases, we used the sample average (SAPS $n=3$; SAPS $n=1$; BPRS $n=2$; NIMH Global Scale Depression $n=1$).

Groups were comparable in terms of sex (COS=56% male; AD=66% male; $\chi^2=0.88$, $df=1$, $p=0.35$), socioeconomic status (based on the Hollingshead and Redlich 1958 H-R Index; $\chi^2=3.1$, $df=4$, $p=0.55$), and illness duration at screening (COS mean=3.12, SD=2.03; AD mean=3.8, SD=2.48; $t=1.78$; $df=136$; $p=0.08$), but were racially distinct (COS=52% Caucasian, 29% African American, 19% Other; AD=79% Caucasian, 13% African American, 8% Other; $\chi^2=10.52$, $df=2$, $p=0.005$). Because the AD group was younger at screening, had an earlier age of psychosis onset, and had a higher IQ score (see Table 1 for means, standard deviations, and t test results), we included these variables in our predictive models.

TABLE 1. GROUP MEANS, STANDARD DEVIATIONS, T TEST RESULTS, UNIVARIATE LOGISTIC REGRESSION COEFFICIENTS AND PREDICTION ACCURACY MEASURES FOR UNIVARIATE AND MULTIPLE LOGISTIC MODELS

	Means (SD)		t test results				Univariate logistic regression results						Multivariate results: 9 predictor model		Multivariate results: 4 predictor model		Multivariate results: 2 predictor model	
			t (136)	p	beta	p	Sensitivity	Specificity	Model accuracy	AUC	Random Forest Importance Z scores ^a	Logistic beta	Random forest importance Z scores ^a	Logistic beta	Random forest importance Z scores ^a	Logistic beta	Random forest importance Z scores ^a	Logistic beta
	COS	AD																
NIMH Global Score:	8.56 (2.55)	5.28 (2.91)	7.51	<0.0001	0.53	0.0001	73.10%	64.94%	71.06%	81.19%	1.65	0.39	0.8	0.49	0.70	0.571		
Psychosis																		
NIMH Global Score:	0.85 (1.26)	2.42 (1.66)	-5.4	<0.0001	-0.61	0.002	73.94%	59.06%	70.22%	74.20%	0.3	-0.67	-0.32	-0.68	-0.70	-0.651		
Depression																		
Total SANS	49.40 (23.86)	24.60 (24.97)	6.54	<0.0001	0.06	0.0003	71.75%	74.63%	72.47%	79.83%	1.37	0.05	0.79	0.04				
Total SAPS	41.08 (18.42)	26.34 (16.20)	5.22	<0.0001	0.06	0.003	68.60%	71.19%	69.25%	74.96%	0.51	0.03	-1.27	0.003				
GAS	31.56 (8.73)	39.17 (9.90)	-4.68	<0.0001	-0.09	0.005	63.65%	66.44%	64.34%	73.08%	-0.88	0.02						
Age at psychosis onset	9.92 (2.06)	8.33 (2.35)	4.13	0.0001	0.32	0.01	65.46%	58.13%	63.63%	67.51%	-0.48	0.35						
Total BPRS	63.87 (13.59)	54.06 (16.95)	3.84	0.0002	0.06	0.02	67.77%	64.00%	66.83%	70.21%	-0.84	-0.01						
Full Scale IQ	71.20 (8.25)	82.00 (16.01)	-3.66	0.0004	-0.04	-0.03	72.21%	64.56%	70.30%	69.18%	-0.79	-0.02						
Atypical	1.12 (0.66)	0.79 (0.69)	2.91	0.004	0.86	0.06	77.48%	35.94%	67.09%	61.91%	-0.84	0.75						
Antipsychotics:																		
count																		
NIMH Global Score:	2.51 (2.47)	3.62 (2.29)	-2.58	0.015	-0.18	0.13	62.52%	53.88%	60.36%	63.23%								
Anxiety																		
Age at screening	13.21 (2.42)	12.29 (2.52)	2.13	0.03														
Illness Duration	3.12 (1.74)	3.80 (2.58)	-1.78	0.08	0.16	0.15	59.92%	53.94%	58.42%	60.06%								
Typical antipsychotics	0.27 (0.46)	0.45 (0.74)	-1.75	0.08														
count																		
NIMH Global Score:	0.93 (1.43)	1.17 (1.18)	-1.06	0.29														
Mania																		
Premorbid problems: ^b	3.87 (2.59)	4.00 (2.90)	-0.27	0.79														
count																		

^aMean decrease in accuracy Z score (score-mean)/SD).

^bAcademic, language, motor, social problems from chart review.

COS, childhood-onset schizophrenia; AD, alternate diagnosis; AUC, area under the curve; NIMH, National Institute for Mental Health; SANS, scale for the assessment of negative symptoms; SAPS, scale for the assessment of positive symptoms; GAS, global assessment of functioning scale; BPRS, brief psychiatric rating scale; IQ, intelligence quotient.

The diagnostic heterogeneity within the AD group encompassed anxiety disorders ($n=19$), mood disorders ($n=12$), pervasive developmental delay disorders ($n=11$), psychotic disorders other than schizophrenia (e.g., psychosis not otherwise specified [NOS]) ($n=24$), eating disorder ($n=1$), and behavioral disorders ($n=25$). Thirty-one of the 53 AD participants had more than one diagnosis. Although notable, the diagnostic heterogeneity in the AD group was necessary for the current study, as psychosis diffuses across diagnostic boundaries, and the purpose of the study was to classify COS versus not COS.

Statistical methods

We began with the following 15 predictors listed in Table 1: age at screening; age at first psychosis; number of typical antipsychotic medications at screening; number of atypical antipsychotic medications at screening; IQ; NIMH Global Scale psychosis, mania, depression, and anxiety ratings; SAPS; SANS; BPRS; GAS; duration of illness at screening; and total number of pre-psychosis/pre-prodrome developmental language, academic, motor, and social problems. We used two sample t tests to determine significant group differences. We also used Wilcoxon rank sum tests to compare groups (not reported), as many of the variables could be considered ordinal: results were not meaningfully different from t test results. We inspected all variables for outliers.

We used statistical (logistic regression) and machine learning (random forest [RF]) methods to examine robustness of results across methods.

Logistic regressions. We selected the 11 predictors that were significantly different between the groups (false discovery rate $q=0.05$) and conducted 11 univariate logistic regressions (dependent variable = group; levels = COS = 1, AD = 0), and corresponding receiver operating characteristic (ROC) analyses using probabilities/fitted values generated by the regression where:

$$\text{Prob(COS)} = P = \frac{e^{\text{logit}}}{1 + e^{\text{logit}}} \quad (1)$$

where e is the base of the natural logarithm and where logit is the log of the odds of being COS estimated by a linear combination of all predictors multiplied by their corresponding estimated coefficients (see Equation 2).

$$\text{logit} = \log_e \left(\frac{P}{1-P} \right) = \log_e(\text{odds}) = b_0 + b_1(x_1) + b_2(x_2) + \dots + b_m(x_m) \quad (2)$$

where m = number of predictors.

Pearson correlations between the 11 predictor variables ranged from 0.001 to 0.72, indicating that multicollinearity among predictors would not be a problem.

Next, we included the nine variables with significant or trend level ($p < 0.10$) univariate logistic regression coefficients in a multiple logistic regression and corresponding ROC analysis (based on probabilities/fitted values generated by the regression). To determine the accuracy of a more concise model, we selected a subset of the four strongest predictors based on univariate results (area under the curve [AUC] > 0.70) and RF variable importance scores (Z score > 0). The four predictors included measures of positive and negative symptoms as well as mood symptoms (SAPS, SANS, NIMH Global Scale Psychosis and Depression scores). Finally, we further reduced the four predictor model by retaining the two items that took the least time to

complete and would be the most practical in a clinical setting (NIMH Global Scale Psychosis and Depression scores). Although a stepwise, forward or backward selection approach could have been applied, there are convincing arguments against such selection methods (Derksen and Keselman 1992; Harrell 2001; Wiegand 2010).

We applied a cross-validation method to estimate model accuracy measures and parameters for all logistic models. Accordingly, for each logistic/ROC model (11 univariate and 3 multiple predictor models), we randomly selected 100 balanced bootstrap (without replacement) samples of 37 participants per group (70% of the smallest group [$n=53$]) to create training sets. For each training set, a corresponding test set was created using those participants who were not selected for the training set. We then averaged results from the 100 test sets to calculate estimates of sensitivity [TP/(TP + FN)], specificity [TN/(TN + FP)], overall accuracy [(sensitivity + specificity)/2], and AUC. For multiple predictor models, we also reported average positive predictive value (PPV) [TP/(TP + FP)], and negative predictive values (NPV) [TN/(TN + FN)] (TP = true positive; TN = true negative; FP = false positive; FN = false negative). For all models, we reported average training set regression coefficients and p values.

Regarding the nine predictor logistic regression, although our events per variables (EPV) ratio was low, there is evidence to support the use of < 10 EPV, as confidence interval coverage is good and bootstrapping methods can reduce bias and validate inferences with < 10 predictors per case (Vittinghoff and McCulloch 2007). We initially considered using all 11 variables associated with a significant group difference in a logistic regression, but because of model convergence problems in the context of a low EPV, we did not pursue this option.

RF. Because logistic regression entails assumptions such as linearity (i.e., the logit link function of the dependent measure is a linear combination of the independent variables) and entails heuristics for the ratio of number of predictors to participants (Harrell et al. 1985; Peduzzi et al. 1996; Vittinghoff and McCulloch 2007), to determine the robustness of our findings across methods and confirm our multiple logistic models, we also conducted RF analyses (Breiman 2001). RF is a popular and straightforward machine learning approach that does not require distributional or model assumptions, and does not have requirements for the ratio of number of predictors to participants. Briefly, RF's basic unit is a classification tree. RF works by selecting a random bootstrap sample per tree (the bagged sample), and at each node in the tree, RF randomly selects a subset of all predictors and selects the variable that best splits data into two daughter nodes. RF determines prediction error by sending the out of bag sample (the part of the sample not selected to grow the tree) down a tree after it is grown. Through this process of selecting bootstrap samples to build the tree and then using the out of bag sample to determine error and variable importance, RF contains an internal cross-validation step, which minimizes overfitting. RF determines group membership by majority vote among the trees when a participant is out of bag: predicted group membership is compared with true group membership to determine prediction accuracy/error. RF also provides importance scores that quantify the magnitude of a variable's effect on accurate predictions.

To limit prediction bias toward the larger group, RF permits sampling the smallest group n per group to build each tree via a tuning parameter, which we set at 53. Other RF parameters included $mtry = 0.5 \times$ number of parameters, number of trees = 1000, terminal node size = 8. We conducted RF-based ROC analyses using probabilities generated from RF run in regression mode and setting COS = 1 and RO = 0 (Malley et al. 2011).

We conducted all analyses in R using the base package, ROCR (Sing et al. 2005) and randomForest (Liaw and Wiener 2002).

TABLE 2. MODEL ACCURACY MEASURES FROM NINE, FOUR AND TWO PREDICTOR MODELS

	Positive predictive value	Negative predictive value	Sensitivity	Specificity	Model accuracy	AUC
9 predictors						
Logistic regression	90.91%	53.84%	77.60%	76.31%	77.28%	86.69%
Random forest	83.75%	86.21%	78.82%	75.47%	77.54%	86.50%
4 predictors						
Logistic regression	91.30%	56.49%	79.69%	76.88%	78.98%	88.15%
Random forest	85.18%	71.93%	81.18%	77.36%	79.71%	86.00%
2 predictors						
Logistic regression	91.43%	55.20%	78.71%	77.56%	78.42%	87.12%
Random forest	80.95%	68.52%	80.00%	69.81%	76.09%	84.00%

AUC, area under the curve.

Results

Group differences

As seen in Table 1, the COS group had higher scores on measures of positive and negative symptoms (SANS, SAPS, BRPS, NIMH Global Scale Psychosis score), lower IQ, worse overall functioning (GAS), and lower NIMH Global Scale anxiety and depression scores. The COS group was older at screening and psychosis onset and had been prescribed a greater number of atypical antipsychotic medications at screening.

Univariate logistic regression results

Each of the 11 variables for which there was a significant group difference was included as a predictor in a univariate logistic regression model. The 11 models ranged from near chance prediction (age at screening, NIMH Global Scale Anxiety score) to 72.47% total accuracy (SANS) with AUC ranging from 60.06% (age at screening) to 81.19% (NIMH Global Scale Psychosis score) (see Table 1).

Multiple logistic regression results

Nine predictor model. We combined the nine variables with either a significant univariate logistic coefficient or a trend ($p < 0.10$) in one additive model. This model performed well, with PPV = 90.91%, NPV = 53.84%, sensitivity = 77.6%, specificity = 76.31%, overall accuracy = 77.28%, and AUC = 86.69%. All 100 bootstrap iterations converged (see Table 2).

Four predictor model. We selected four predictors that had RF importance Z scores > 0 and univariate AUC > 70 (see Fig. 1): SAPS, SANS, and NIMH Global Scale Psychosis and Depression scores. The four predictor logistic model had PPV = 91.30%, NPV = 56.49%, sensitivity = 79.69%, specificity = 76.88%, overall accuracy = 78.98%, and AUC = 88.15% (see Table 2).

Two predictor model. We further reduced the model to include only NIMH Global Scale psychosis and depression scores. These features were selected because they are the easiest to complete and the most practical in a clinical setting. The two predictor model had accuracy measures comparable to the four predictor model: PPV = 91.343%, NPV = 55.20%, sensitivity = 78.71%, specificity = 77.56%, overall accuracy = 78.42%, and AUC = 87.12% (see Table 2).

RF results

The nine predictor RF analysis had PPV = 83.75, NPV = 86.12%, sensitivity = 78.82%, specificity = 75.47%, overall accuracy = 77.54%,

and AUC = 86.50%. The four predictor RF analysis had PPV = 85.18%, NPV = 71.93%, sensitivity = 81.18%, specificity = 77.36%, overall accuracy = 79.71%, and AUC = 86%. The two predictor RF analysis had PPV = 80.95%, NPV = 68.52%, sensitivity = 80%, specificity = 69.81%, overall accuracy = 76.09%, and AUC = 84% (see Table 2). Variable importance scores tracked with univariate AUC scores ($r = 0.91$ for univariate AUC and nine variable RF); see Table 1 for variable importance scores from the nine, four, and two predictor RF analyses.

In summary, the results show that COS and AD groups can be separated with good sensitivity, specificity, and PPV using either logistic regression or RF. NPV is worse ($\sim 50\%$) with logistic regression.

Discussion

Using information available at screening, we predicted COS status after a lengthy inpatient hospitalization that included a medication washout. We tested models with nine, four, and two predictors and they all performed well. Results were also robust across methods.

FIG. 1. Selecting four predictor model based on univariate area under the curve (AUC) and random forest importance scores.

Notably, using only two predictors in a logistic regression—NIMH Global Scale psychosis and depression scores—we were able to classify COS versus impaired children and adolescents with non-fleeting, non-trivial levels of psychosis with 91% PPV (chance that a patient is COS given that the model predicts they are COS), 55% NPV, 79% sensitivity, 78% specificity, and 78% overall accuracy. This model performed essentially as well as the larger logistic models, indicating that it can be efficiently applied without losing a meaningful amount of prediction accuracy. The results from the models indicate that higher psychosis ratings and lower depression ratings combine to increase the probability that a patient has COS.

The NIMH Global Scale depression and psychosis ratings are each a single item scored on a scale of 1–15, with higher scores reflecting greater severity (see Appendix 1). They are completed after thorough clinical interviews with patients and parents conducted by trained psychiatrists and social workers. As such, clinicians can quantify their clinical observations and enter the two scores in Appendix 2, which is an Excel worksheet, to obtain the probability that a patient has COS. Using a cutoff of 50%, if the calculated probability of COS in the Appendix 2 worksheet is >50%, the probability that the patient has COS is 91% (PPV). For example, given an NIMH Global Scale Psychosis score of 12 and NIMH Global Scale Depression score of 4, a patient has a 79.76% probability of having COS. This increases to 96.52% if the NIMH Global Scale Depression score drops to 1. In both cases, these probabilities (79.76% and 96.52%) can be used to make a dichotomous diagnostic classification of COS versus not COS: with the cutoff of 50% (>0.50, a patient is classified as COS), both example patients would be classified as having COS with a PPV of 91%. At the same time, users must bear in mind that, given a predicated probability <0.50, there is a 55% chance a patient has an alternate diagnosis (NPV).

Because there have been successful efforts classifying schizophrenia patients and controls using machine learning methods and magnetic resonance imaging (MRI) measures (Davatzikos et al. 2005; Kawasaki et al. 2007; Yoon et al. 2007; Koutsouleris et al. 2009; Sun et al. 2009), we explored using structural MRI measures as potential variables to include in our classification schemes. Using only regional measures of cortical thickness yielded weak prediction (e.g., 65% accuracy), suggesting that clinical information may have more power than MRI measures when classifying COS versus non-fleeting, non-trivial levels of psychosis.

In the past several years, the increased emphasis on a dimensional conceptualization of psychosis with respect to childhood psychiatric illness has been an important advance in understanding childhood psychopathology (Polanczyk et al. 2010). This article is not meant to contrast with advances in this area. On the contrary, we wish to use the dimensionality evident in psychosis to help clinicians for whom a diagnosis is important for treatment decisions and outcome expectations.

On a technical note, the differences between RF and logistic regression models' sensitivity, specificity, AUC, and PPV results were minimal. Also, RF importance scores results tracked with results from univariate statistical models, consistent with our previous work (Greenstein et al. 2012), and RF probabilities were highly correlated with logistic probabilities (not reported). These consistencies suggest that our logistic models met statistical assumptions and contain minimal bias. At the same time, all multiple logistic models have stronger PPV than NPV. This is the case when false negatives are relatively more frequent than false positives and/or true positives are relatively more frequent than true negative. This can be true when there are more cases (COS) than non-cases (AD), as in our sample. As such, our logistic models tend to favor misclassifying COS as AD over misclassifying AD as COS, making the likelihood of COS

quite high if a patient is classified as having COS. In contrast, if a patient is classified as having AD using any of the logistic models (two, four and nine predictor models), that patient may or may not belong in the AD group, with almost equal chances. In contrast to the multiple logistic models, the RF PPVs and NPVs are more consistent.

Limitations of the current study include the fact that our best false negative rate (COS classified as AD) (18.82%) is far worse than our 20 year false negative rate/misdiagnoses of COS as AD: only four AD participants were later diagnosed with schizophrenia at follow-up. This is not surprising, and is most likely because our inpatient washout admission criteria are intentionally biased to avoid false negatives/missing a COS case.

The calculated probabilities from the Appendix 2 worksheet must be interpreted with caution, given that the generalizability of our sample is limited to children and adolescents who have a marked history of reported psychosis before the age of 13, a concomitant deterioration in functioning, absence of neurological disease, and a Full Scale IQ score > 70. We may also have a sampling bias involving severely impaired patients with relatively healthy families and parents who are capable of applying for a 2–5 month inpatient hospitalization, often far from home. An additional limitation is that raters in the community and raters in our study are not necessarily reliable with each other. Lastly, the NIMH Global Scale psychosis and depression scales are completed in the context of the remainder of our ratings, whereas the appendix we are attaching will presumably be completed in isolation, additionally limiting generalization. Given these potential limits to the generalizability of our algorithm, we encourage users to report on the accuracy of the algorithm in other settings.

Conclusions

We have condensed and quantified >20 years of experience to develop diagnostic classification schemes that, we hope, will be of use to clinicians. The logistic regression classifier we include as a worksheet has a sensitivity of 78.71%, a specificity of 77.56%, a PPV of 91%, and an NPV of 55%. Results should be interpreted in light of generalizability limitations.

Clinical Significance

Based on the analyses conducted in this study, we are providing a worksheet that clinicians can use to determine the likelihood a child or adolescent has COS. Using information from a clinical interview, a clinician can complete the ratings included in Appendix 1 and enter the scores in a worksheet (see Appendix 2). When the clinician enters the scores, the worksheet calculates the probability that the patient has COS. This number, along with other relevant information can be used to help guide treatment and diagnostic decisions. This worksheet should only be used for child and adolescent patients who have psychosis onset before the age of 13.

Disclosures

No competing financial interests exist.

References

- Andreasen NC: Negative symptoms in schizophrenia: Definition and reliability. *Arch Gen Psychiatry* 39:784–788, 1982.
- Andreasen NC: Scale for the Assessment of Negative Symptoms (SANS). Iowa City: University of Iowa; 1983.
- Andreasen NC: Scale for the Assessment of Positive Symptoms (SAPS). Iowa City: University of Iowa; 1984.
- Asarnow JR, Benmeir S: Children with schizophrenia spectrum and depressive-disorders – a comparative-study of premorbid adjustment,

- onset pattern and severity of impairment. *J Child Psychol Psychiatry* 29:477–488, 1988.
- American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. Washington, DC: American Psychiatric Association; 1994.
- Bartels–Velthuis AA, van de Willige G, Jenner JA, van Os J, Wiersma D: Course of auditory vocal hallucinations in childhood: 5-year follow-up study. *Br J Psychiatry* 199:296–302, 2011.
- Breiman L: Random forests. *Mach Learn* 45:5–32, 2001.
- Bunney WE, Hamburg DA: Methods for reliable longitudinal observation of behavior. *Arch Gen Psychiatry* 9:280–294, 1963.
- Davatzikos C, Shen D, Gur RC, Wu X, Liu D, Fan Y, Hughett, P., Turetsky BI, Gur, RE: Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. *Arch Gen Psychiatry* 62:1218–1227, 2005.
- Derksen S, Keselman HJ: Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *Br J Math Statist Psychol* 45:265–282, 1992.
- Gochman P, Miller R, Rapoport JL: Childhood-onset schizophrenia: The challenge of diagnosis. *Curr Psychiatry Rep* 13:321–322, 2011.
- Greenstein D, Malley JD, Weisinger B, Clasen L, Gogtay N: Using multivariate machine learning methods and structural MRI to classify childhood onset schizophrenia and healthy controls. *Front Psychiatry* 3:53, 2012.
- Hafkenscheid A: Reliability of a standardized and expanded Brief Psychiatric Rating Scale: A replication study. *Acta Psychiatr Scand* 88:305–310, 1993.
- Harrell F: *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer; 2001.
- Harrell FE, Lee KL, Matchar DB, Reichert TA: Regression models for prognostic prediction: Advantages, problems, and suggested solutions. *Cancer Treat Rep* 69: 1071–1077, 1985.
- Hollis C: Child and adolescent (juvenile onset) schizophrenia: A case control study of premorbid developmental impairments. *Br J Psychiatry* 166:489–495, 1995.
- Horwood J, Salvi G, Thomas K, Duffy L, Gunnell D, Hollis C, Lewis G, Menezes P, Thompson A, Wolke D, Zammit S, Harrison G: IQ and non-clinical psychotic symptoms in 12-year-olds: Results from the ALSPAC birth cohort. *Br J Psychiatry* 193:185–191, 2008.
- Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N: Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 37:980–988, 1997.
- Kawasaki Y, Suzuki M, Kherif F, Takahashi T, Zhou S, Nakamura K, Matsui M, Sumiyoshi T, Seto H, Kurachi M: Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. *NeuroImage* 34:235–242, 2007.
- Koutsouleris N, Meisenzahl EM, Davatzikos C, Bottlender R, Frodl T, Scheuerecker J, Schmitt G, Zetzsche T, Decker P, Reiser M, Möller HJ, Gaser, Christian: Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch Gen Psychiatry* 66:700–712, 2009.
- Liaw A, Wiener M: Classification and Regression by random Forest. *R News* 2:18–22, 2002.
- Lukoff D, Nuechterlein KH, Ventura J: Manual for the expanded Brief Psychiatric Rating Scale (BPRS) Schizophrenia Bulletin. *Schizophr Bull* 12:594–602, 1986.
- Malley JD, Kruppa J, Dasgupta A, Malley K, Ziegler A: Probability machines. consistent probability estimation using nonparametric learning machines. *Methods Inf Med* 50:74–81, 2011.
- McKenna K, Gordon CT, Lenane M, Kaysen D, Fahey K, Rapoport J: Looking for childhood onset schizophrenia: The first 71 cases screened. *J Am Acad Child Adolesc Psychiatry* 33:636–644, 1994.
- Murphy DL, Pickar D, Alterman IS: Methods for the quantitative assessment of depressive and manic behavior. In: *The Behavior of Psychiatric Patients*. Edited by E.L. Burdock, A. Sudilovsky, S. Gershon. New York: Marcel Dekker, 355–392, 1982.
- Nicolson R, Lenane M, Brookner F, Gochman P, Kumra S, Spechler L, Giedd JN, Thakar GK, Wudarsky M, Rapoport JL: Children and adolescents with psychotic disorder not otherwise specified: A 2- to 8-year follow-up study. *Compr Psychiatry* 42:319–325, 2001.
- Nicolson R, Rapoport JL: Childhood-onset schizophrenia: rare but worth studying. *Biol. Psychiatry* 46:1418–1428, 1999.
- Norman RMG, Malla AK, Cortese L, Diaz F: A study of the inter-relationship between and comparative interrater reliability of the SAPS, SANS and PANSS. *Schizophr. Res.* 19:73–85, 1996.
- Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 49: 1373–1379, 1996.
- Polanczyk G, Moffitt TE, Arseneault L, Cannon M, Ambler A, Keefe RS, Houts R, Odgers CL, Caspi, A: Etiological and clinical features of childhood psychotic symptoms: results from a birth cohort. *Arch Gen Psychiatry* 67:328–338, 2010.
- Poulton R, Caspi A, Moffitt TE, Cannon M, Murray R, Harrington H: Children's self-reported psychotic symptoms and adult schizophreniform disorder: A 15-year longitudinal study. *Arch Gen Psychiatry* 57:1053–1058, 2000.
- Sing T, Sander O, Beerenwinkel N, Lengauer T: ROCRC: Visualizing classifier performance in R. *Bioinformatics* 21:3940–3941, 2005.
- Stayer C, Sporn A, Gogtay N, Tossell JW, Lenane M, Gochman P, Greenstein D, Sharp W, Rapoport JL: Multidimensionally impaired: the good news. *J Child Adolesc Psychopharmacol* 15:510–519, 2005.
- Sun D, van Erp T, Thompson P, Bearden CE, Daley MK, Kushan L, Hardt ME, Nuechterlein KH, Toga AW, Cannon TD: Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: Classification analysis using probabilistic brain atlas and machine learning algorithms. *Biol. Psychiatry* 66:1055–1060, 2009.
- Sunderland T, Alterman IS, Yount D, Hill JL, Tariot PN, Newhouse PA, Mueller EA, Mellow AM, Cohen RM: A new scale for the assessment of depressed mood in demented patients. *Am J Psychiatry* 145:955–959, 1988.
- van Os J, Linscott RJ, Myin–Germeys I, Delespaul P, Krabbendam L: A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. *Psychol Med* 39:179–195, 2009.
- Vittinghoff E, McCulloch CE: Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol* 165:710–718, 2007.
- Wiegand RE: Performance of using multiple stepwise algorithms for variable selection. *Stat Med* 29:1647–1659, 2010.
- Yoon U, Lee JM, Im K, Shin YW, Cho BW, Kim IY, Kwon JS, Kim SI: Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. *NeuroImage* 34:1405–1415, 2007.

Address correspondence to:
Deanna Greenstein, PhD
10 Center Drive Room 3N202
Bethesda, MD 20892

E-mail: dede.greenstein@nih.gov

(Appendix follows →)

APPENDIX 1. NIMH GLOBAL SCALE: RATINGS FOR PSYCHOSIS AND DEPRESSION*

<i>Scale</i>	<i>Psychosis</i>	<i>Depression</i>
1–3 Minimal	Ratings of 2–3 may reflect odd or strange manner, apathy, somewhat flat affect, social withdrawal, inattention or suspiciousness.	Ratings of 2–3 reflect some sadness, gloominess, pessimism, sluggishness, or mildly diminished interests, or diminished sense of competence.
4–6 Mild	Including some distortions of reality, difficulties with logic, instances of inappropriate affect, or inappropriate interpersonal relations. May rarely report hearing voices without responding to them or noninterfering delusions, without elaboration.	More persistent depressive symptoms which may be directly expressed as sadness and depressive feelings, or as other symptoms such as somatic complaints, some feelings of inability to cope, increased dependency, disinterest in usual activities, or feeling slowed down and less able to function. Ordinarily, such symptoms would be somewhat evident to friends and relatives.
7–9 Moderate	With more evident conceptual disorganization, some reality testing and contact with reality is maintained (e.g. patient will consider staff explanations), but patient could not function for more than a day or two without hospitalization. Symptoms may include hallucinations, throughout the day, interfering delusions, severe thought blocking, loose associations, markedly inappropriate affect.	More pervasive depressive feelings, with helpless and hopelessness, social withdrawal, some psychomotor retardation and/or anxious agitation, and often problems sleeping (e.g. early morning awakening or excess sleep). Symptoms would usually lead to seeking treatment including hospitalization.
10–12 Severe	Major loss of contact with reality. Multiple psychotic symptoms are present including definite thought disorder, pervasive involvement with hallucinations, preoccupation with very bizarre thoughts or ideas, little control over behavior, and inability to function outside a hospital.	Depressive symptoms as described above, but associated with more marked helplessness, a sense of worthlessness, and often preoccupation with death. There may be impaired judgment or loss of insight, non-communicativeness, and more retardation and/or agitation. Patient is unable to function outside a hospital.
13–15 Very Severe	Absence of reality contact, with loss of ego boundaries. Multiple symptoms of psychosis continually present, although patient will often be sufficiently out of touch that he may not verbalize symptoms. Frequently characterized by catatonia, severe agitation or combativeness, smearing, or word salad. Minimal self-care is usually impossible for the patient.	Depressive symptoms as above, but symptoms (e.g. muteness, extreme agitation, or retardation) are more profound and incapacitating. May also include depressive delusions (somatic delusions or delusions of guilt) or regressed behavior. Patients may require close supervision for eating and other basic activities.

*Murphy DL, Pickar D, Alterman IS: Methods for the quantitative assessment of depressive and manic behavior. In: The Behavior of Psychiatric Patients. Edited by Burdock EL, Sudilovsky A, Gershon S. New York, Marcel Dekker, 1982, pp 355–392.

*Sunderland T, Alterman IS, Yount D, Hill JL, Tariot PN, Newhouse PA, Mueller EA, Mellow AM, Cohen RM. A new scale for the assessment of depressed mood in demented patients. Am J Psychiatry 145(8); 955–959, 1988.

APPENDIX 2. NIMH GLOBAL SCALE DEPRESSION AND PSYCHOSIS RATINGS WORKSHEET

	<i>Enter score</i>
NIMH Global Scale Psychosis Score ^{a,b}	12
NIMH Global Scale Depression Score ^{a,b}	4
Probability of childhood-onset schizophrenia	0.7976

^aNIMH Global Scale (Murphy DL, Pickar D, Alterman IS: Methods for the quantitative assessment of depressive and manic behavior. In: The Behavior of Psychiatric Patients. Edited by Burdock EL, Sudilovsky A, Gershon S. New York, Marcel Dekker, 1982, pp 355–392).

^bSunderland T, Alterman IS, Yount D, Hill JL, Tariot PN, Newhouse PA, Mueller EA, Mellow AM, Cohen RM. A new scale for the assessment of depressed mood in demented patients. Am J Psychiatry 145:955–959, 1988.