Neurocognitive Pattern Analysis Reveals Classificatory Hierarchy of Attention Deficits in Schizophrenia

Christina Shen\(^1\), Florin C. Popescu\(^2\), Eric Hahn\(^1\), Tam T.M. Ta\(^1\), Michael Dettling\(^1\), and Andres H. Neuhaus\(^*\,\(^1\)

\(^1\)Department of Psychiatry and Psychotherapy, Charité University Medicine, Berlin, Germany; \(^2\)Fraunhofer Institute for Open Communication Systems FOKUS, Berlin, Germany

*To whom correspondence should be addressed; Department of Psychiatry and Psychotherapy, Charité University Medicine, Campus Benjamin Franklin, Eschenallee 3, 14050 Berlin, Germany; tel: 49-30-8445-8412, fax: 49-30-8445-8393; e-mail: andres.neuhaus@charite.de

Attention deficits, among other cognitive deficits, are frequently observed in schizophrenia. Although valid and reliable neurocognitive tasks have been established to assess attention deficits in schizophrenia, the hierarchical value of those tests as diagnostic discriminants on a single-subject level remains unclear. Thus, much research is devoted to attention deficits that are unlikely to be translated into clinical practice. On the other hand, a clear hierarchy of attention deficits in schizophrenia could considerably aid diagnostic decisions and may prove beneficial for longitudinal monitoring of therapeutic advances. To propose a diagnostic hierarchy of attention deficits in schizophrenia, we investigated several facets of attention in 86 schizophrenia patients and 86 healthy controls using a set of established attention tests. We applied state-of-the-art machine learning algorithms to determine attentive test variables that enable an automated differentiation between schizophrenia patients and healthy controls. After feature preranking, hypothesis building, and hypothesis validation, the polynomial support vector machine classifier achieved a classification accuracy of 90.70% ± 2.9% using psychomotor speed and 3 different attention parameters derived from sustained and divided attention tasks. Our study proposes, to the best of our knowledge, the first hierarchy of attention deficits in schizophrenia by identifying the most discriminative attention parameters among a variety of attention deficits found in schizophrenia patients. Our results offer a starting point for hierarchy building of schizophrenia-associated attention deficits and contribute to translating these concepts into diagnostic and therapeutic practice on a single-subject level.

Key words: focused attention/sustained attention/ selective attention/alternating attention/ divided attention/machine learning

Introduction

The identification of biological markers of schizophrenia is a central desiderate of biological psychiatry. However, reliable and objective biomarkers are still missing and clinicians depend on formal diagnostic criteria and their clinical impression, consequently running the risk to bias objective criteria toward subjective experience.

Cognitive deficits are among the most consistent findings in schizophrenia research and represent an important predictor of functional outcome.\(^1,\(^2\)\) Meta-analyses have suggested generalized deficits of various cognitive functions in schizophrenia, including verbal and nonverbal memory, motor performance, attention, general intelligence, and executive function.\(^3,\(^4\)\) Notably, high heterogeneity and discrepancy among the studies are reported.\(^5,\(^6\)\) Several neurocognitive test batteries, such as the Measurement and Treatment Research to Improve Cognition in Schizophrenia, the Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia, or the Cambridge Neuropsychological Test Automated Battery, have been established, containing valid and reliable neurocognitive tasks in order to assess deficits of different cognitive domains in schizophrenia and cognitive effects of compounds in clinical trials.\(^7,\(^8\)\) However, given the nonspecific nature of cognitive deficits in schizophrenia and the heterogeneity of findings across studies, the value of those tests as diagnostic differentiators between schizophrenia and healthy subjects remains unclear.

Among the variety of cognitive deficits in schizophrenia, deficits of attention are most robust findings.\(^10,\(^11\)\) Attention is usually fractioned into functional subcomponents;\(^12,\(^14\)\) accordingly, neurocognitive schizophrenia research usually focuses on specific attention subcomponents that are assessed by numerous different tasks. For example, deficits of sustained attention have been frequently
observed in different versions of the Continuous Performance Test (CPT).\textsuperscript{15-17} selective attention deficits involving alerting, orienting, and conflict were found using the Attention Network Test (ANT).\textsuperscript{18,19} reduced performance was further observed in variants of the Wisconsin Card Sorting Test (WCST), the Stroop task, Trail Making test, or masking tasks.\textsuperscript{20-25} Complicating reconciliation of these studies, performance deficits are often assigned to different cognitive functions across studies and/or to a diffuse nosological terminology that does not clearly distinguish between, eg, “executive,” “prefrontal,” or “top-down,” functions. In this context, agreeing upon a commonly accepted model of attention could considerably aid diagnostic decisions and may prove beneficial for longitudinal monitoring of therapeutic advances.

The present study focuses on the clinical model of attention proposed by Sohlberg and Mateer\textsuperscript{13} that distinguishes focused, sustained, divided, alternating, and selective attention on the basis of patient histories who successively recover these attention domains after suffering from organic brain damage. The advantage of using this model for machine learning classification is based on integrating only those attention domains that are clinically clearly separable and thus do not contain overlapping domains, which would inevitably bias any classificatory model toward these overlaps. We employed an evenly distributed set of attention domain tests in a large cohort of schizophrenia patients and healthy controls and applied state-of-the-art machine learning algorithms to determine those attention domain test variables that differentiate between patients and controls on a single-subject level.

Methods

Subjects

Eighty-six schizophrenia patients (38 females, 48 males) participated in this study. They met DSM-IV criteria for schizophrenia, and had no psychiatric disorder other than schizophrenia and nicotine abuse/dependence. Current drug abuse as determined by urine toxicology led to exclusion from the study. None of the included patients had a history of severe medical disorder or severe neurological disorder or history of electroconvulsive therapy. All patients were recruited from the inpatient unit and the outpatient facility of the Department of Psychiatry, Charité University Medicine, Berlin, Germany.

Eighty-six healthy subjects (34 females, 52 males) recruited via newspaper served as controls. Control participants were screened for mental and physical health and were excluded when fulfilling the criteria of psychiatric disorders according to DSM-IV as determined by structured clinical interviews.\textsuperscript{26} Further reasons for exclusion were a family history of psychiatric illness, medical or neurological disorders, or intake of psychotropic drugs as confirmed by urine toxicology before participating in the study.

All participants were right handed and reported normal or corrected-to-normal vision and audition. All subjects gave written informed consent before participating in this study. This study was approved by the ethics committee of the University Hospital Benjamin Franklin, Charité University Medicine, Berlin, Germany, and was conducted in accordance with the declaration of Helsinki and its subsequent amendments.

Cognitive Tasks

All participants completed a series of hierarchically organized attention domain tests including Continuous Performance Test-Identical Pairs (CPT-IP), ANT, WCST, and alerting and divided attention subtests of a German test of attention performance (Testbatterie zur Aufmerksamkeitsprüfung; TAP; version 1.7). The TAP contains 13 subtests evaluating different components of attention.\textsuperscript{27} In our study, participants completed subtests for focused and divided attention. Every attention test was performed on a personal computer with a 17-inch cathode ray tube monitor. For every test, a set of 6 variables was extracted for further analyses.

Focused Attention. In this TAP subtest, participants were instructed to press the response key whenever a cross appears on the screen. The appearance of a cross was preceded by a warning tone. Following a training session with 5 trials, a total of 40 test trials were completed. Outcome measures consisted of mean reaction time (RT), number of valid reactions, anticipation errors, errors of omission, lapses of attention, and a specific parameter indicative of phasic alertness.

Sustained Attention. The CPT-IP was developed as a test of sustained visual attention in both schizophrenia patients and healthy controls.\textsuperscript{28} Using their dominant hand, participants pressed the mouse button as fast as possible when 2 identical pairs of symbols were presented in sequence. Following a trainings session with 10 trials, 150 stimuli were presented with an invariant stimulus-onset asynchrony of 1000 ms and a stimulus duration of 50 ms. Thirty out of 150 stimuli served as target stimuli. Outcome measures were the $d'$ and $\log \beta$, standard measures in signal detection theory representing the signal-to-noise ratio ($d'$), and the ratio of signal likelihood to noise likelihood ($\log \beta$); hit rate (percentage correct hits) and mean RT for correct hits; and percentage of and RT for false alarms.

Selective Attention. The ANT is a test of selective visual attention that combines a cued detection paradigm with a flanker task.\textsuperscript{29} In the task, informative cues are presented for 100 ms, followed by targets with a stimulus-onset
asynchrony of 500 ms. The cues appeared in the center, superimposed on the fixation cross (center cue; only temporally informative); simultaneously above and below the center (double cue; only temporally informative); either above or below the fixation cross (spatial cue; temporally and spatially informative); or were not presented (no cue). Targets consisted of 5 horizontally arranged arrows or lines presented above or below the fixation cross. Participants were instructed to indicate the direction of a central arrow by making a left or right button press on a keyboard, irrespective of the (compatible or incompatible) flanking arrows. Following a training session of 24 pseudo-randomized trials with feedback, participants completed a total of 288 pseudo-randomized trials, divided into 3 experimental blocks, without feedback. Outcome measures were attention network effects of alerting, orienting, and conflict, calculated as the difference in RT between different task conditions. Other dependent variables extracted for machine learning analysis were mean RT across trials, errors of commission, and errors of omission.

**Alternating Attention.** The WCST measures alternating attention by assessing estimates of establishing and shifting cognitive sets. Participants are instructed to sort stimulus cards on the basis of color, form, or number of symbols. The only feedback provided is whether the current response was correct or incorrect. The sorting rule changed after 10 consecutive correct responses. The test was discontinued when the participant has learned 2 iterations of each sorting rule, or completed 128 trials. The primary outcome measures used for this study were failures to maintain set, numbers of conceptual level responses, perseverative responses, perseverative errors, nonperseverative errors, and number of categories completed.

**Divided Attention.** The TAP subtest for divided attention combines a visual and an auditory task, both of which have to be processed in parallel. During the test, varying numbers of crosses appeared on the screen simultaneously, while a high and a low tone were presented in sequence at the same time. Participants were instructed to press a response key when 4 crosses formed a square or when the same tone was presented twice in a row. Following a training session of 10 trials, a total of 50 test trials was presented. Outcome measures were mean RT, number of valid reactions, errors of commission, errors of omission, lapses of attention, and late reactions.

Additionally, all participants were tested with a basic neuropsychological test battery including verbal and nonverbal intelligence tests along with the Digit Symbol Test and the Trail Making Test (see Table 1).

**Machine Learning Pattern Classification**

The machine learning discovery methods involve 3 separate stages of analysis: feature selection, hypothesis ranking, and validation. The bulk of this analysis was done in Matlab (Mathworks) with help from the open-source packages Weka (data mining software in Java, available online at [http://www.cs.waikato.ac.nz/ml/weka/](http://www.cs.waikato.ac.nz/ml/weka/)), LibSVM (library for support vector machines, available online at [http://www.csie.ntu.edu.tw/~cjlin/libsvm/](http://www.csie.ntu.edu.tw/~cjlin/libsvm/)), and bolasso.m (bootstrap-enhanced least absolute shrinkage operator, available online at [http://code.google.com](http://code.google.com)).

Machine learning analysis was done entering \( n = 30 \) outcome measures of attention tasks. Digit Symbol Test (\( N \) correct) and IQ (mean IQ, averaged across verbal and nonverbal domain) were entered as additional features because of substantial evidence for a large

### Table 1. Summary of Demographic and Clinical Data

<table>
<thead>
<tr>
<th></th>
<th>Schizophrenia</th>
<th>Controls</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N ) (females/males)</td>
<td>86 (34/52)</td>
<td>86 (38/48)</td>
<td>.033</td>
</tr>
<tr>
<td>Age (y)</td>
<td>36.33 ± 10.6</td>
<td>32.58 ± 9.3</td>
<td>.015</td>
</tr>
<tr>
<td>Age (range)</td>
<td>18–67</td>
<td>18–65</td>
<td>—</td>
</tr>
<tr>
<td>Education (y)</td>
<td>13.33 ± 2.6</td>
<td>15.29 ± 2.4</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>106.18 ± 14.8</td>
<td>114.93 ± 14.1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Nonverbal IQ</td>
<td>103.75 ± 11.6</td>
<td>114.29 ± 8.6</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Digit Symbol Test</td>
<td>43.66 ± 11.8</td>
<td>60.07 ± 12.9</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Trail Making Test—A</td>
<td>36.48 ± 15.2</td>
<td>26.14 ± 6.8</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Trail Making Test—B</td>
<td>93.84 ± 51.7</td>
<td>56.62 ± 18.1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>DOI (mo)</td>
<td>97.21 ± 101.8</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>N episodes</td>
<td>3.54 ± 2.8</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>PANSS positive scale</td>
<td>13.18 ± 4.3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>PANSS negative scale</td>
<td>17.46 ± 6.3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>PANSS general scale</td>
<td>31.89 ± 9.1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CPZ equivalents (mg/d)</td>
<td>510.25 ± 374.2</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

**Note:** DOI, duration of illness; PANSS, Positive And Negative Syndrome Scale; CPZ, chlorpromazine.

Between-group differences were assessed with \( t \) test, for independent samples, or chi-square test, as appropriate.
effect schizophrenia diagnosis on both Digit Symbol Test (DST)\(^6\) and IQ performance,\(^14\) so that these measures may add to classification accuracy on a single-subject level.

The first 2 stages involve a randomly chosen subset of the entire data set which comprises 75% of the data, ie, the training subset. In the first stage (feature selection), features (preselected for relevance on physiological grounds) are ranked according to each feature’s strength of difference between control and patient groups as measured by the Kolmogorov-Smirnov test. In order to offer an intuitive measure of strength of difference (other than the P value of the Kolmogorov-Smirnov test significance) we also provide the classification value of the linear discriminant analysis classifier, meaning the division of the real line into 2 segments (the simplest means of classifying a 1 dimensional value). The justification for providing such an error value is to give an approximate lower bound for possible final classification accuracy values of more complex, multidimensional classifiers.

In the next step of analysis (hypothesis ranking), a large number of hypotheses are evaluated over the training subset in terms of their classification error measured according to the 10-fold cross-validated balanced error rate, which, in the case of equal numbers of samples from each class, is the same as the error rate proper. The hypotheses are formed as such: a candidate feature combination set is constructed and consists of exhaustive combinatorial subsets with up to 5 maximal features, to which an element consisting of all strong features is added. A set product of the candidate feature combination set and the classifier set is formed which provides the hypothesis set. The classifier set is a list of common classification algorithms used in machine learning practice, and is composed of: linear and quadratic discriminants (LDA/QDA and their diagonal variants), support vector machines (SVM with radial basis, linear, polynomial, and multilayer perceptron kernels), naive Bayes, k-nearest neighbors with Euclidean and cosine distance measures, and Mahalanobis with different types of covariance estimators (no regularization, robust regression, and regularized/sparsified estimates: model consistent Lasso estimation through the bootstrap [BOLASSO] and minimum description length [MDL]). A total of 15 classifier types is thus estimated, and a total of 405 hypotheses per data set are tested. The testing of a much larger set of hypotheses, far greater than the number of samples, is liable to over-fit even at this stage. The 10-fold cross-validated (optimizing on 9/10 of the training data, evaluating on the remaining 1/10th, repeated 10 times and averaged) classification error called the predicted error is used to a final ranking of hypotheses in the second stage of analysis. Hypotheses whose predicted accuracy was worse than other considered hypotheses with the same set of features were discarded.

In the third and final step (validation), hypotheses are evaluated (best-to-worst) on the validation subset, ie, 25% remaining from the original 75%/25% split, which until this step has not played part in any fitting or optimization. Hypotheses that resulted in tested over-fit according to a 2-sided binomial test were discarded.

Results

Table 2 summarizes the results from feature prerranking according to the accuracy of individual discrimination between controls and patients, based on separately calculated between-group significances with Kolmogorov-Smirnov test. This first step in our machine learning algorithm allowed for exclusion of noninformative cognitive test variables and established a preliminary set of 18 cognitive measures that principally discriminate between groups.

Based on feature prerranking, we subsequently identified strong features that differentiate between control and patient groups. These identified strong features are illustrated in figure 1. This procedure allowed for excluding the majority of variables, thus leaving a set of 5 cognitive measures that are further grouped into hypothetical clusters based on their between-group differences of cumulative probability density function (see figure 1).

Following hypothesis ranking, the top 5 hypotheses identified are shown in table 3, in terms of predicted and validation errors, along with the expected SD of the binomial distribution based on the validation error and the number of samples in the validation set. Note that the predicted and validated errors are within the margin of error. The results revealed validation errors of around 9%–12% along with a mean SD of around 3% among the top 5 hypotheses. Here, radial basis function SVM achieved a classification accuracy of 90.70% ± 2.9% using psychomotor speed and 3 different attention parameters that belonged to the sustained and divided attention domains.

Of note, chosen features were identified by virtually all machines and are not simply the first N-ranked features, but subsets of features.

Discussion

We investigated specific attention domains according to the clinical model of attention proposed by Sohlberg and Mateer\(^13\) in schizophrenia patients and healthy controls. Using advanced machine learning methods, we identified a set of 4 test features that sufficed to obtain a cumulative classification accuracy of 90.70%. The closeness of the prediction error (12.38%) to the validation error (9.30%) is particularly convincing and argues against overfitting the mathematical model to the data. The identified feature set consisted of variables extracted from psychomotor speed as well as sustained and divided attention tasks that distinguish between patients and controls over and above statistical significance. This classification establishes, to the best of our knowledge, the first hierarchy of attention deficits in schizophrenia, suggesting superiority.
Table 2. Preranking of Statistically Significant Test Variables

<table>
<thead>
<tr>
<th>Test variable</th>
<th>$P^a$</th>
<th>Cohen's $d^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DST</td>
<td>&lt;.001</td>
<td>1.33</td>
</tr>
<tr>
<td>CPT-IP: hits</td>
<td>&lt;.001</td>
<td>1.17</td>
</tr>
<tr>
<td>IQ</td>
<td>&lt;.001</td>
<td>1.00</td>
</tr>
<tr>
<td>TAP (divided attention): valid reactions</td>
<td>&lt;.001</td>
<td>0.96</td>
</tr>
<tr>
<td>ANT: mean RT</td>
<td>&lt;.001</td>
<td>0.83</td>
</tr>
<tr>
<td>TAP (divided attention): errors of omission</td>
<td>&lt;.001</td>
<td>0.83</td>
</tr>
<tr>
<td>TAP (divided attention): mean RT</td>
<td>&lt;.001</td>
<td>0.78</td>
</tr>
<tr>
<td>CPT-IP: $d'$</td>
<td>&lt;.001</td>
<td>0.77</td>
</tr>
<tr>
<td>TAP (focused attention): mean RT</td>
<td>&lt;.001</td>
<td>0.72</td>
</tr>
<tr>
<td>WCST: nonperseverative errors</td>
<td>.001</td>
<td>0.71</td>
</tr>
<tr>
<td>WCST: categories completed</td>
<td>.002</td>
<td>0.70</td>
</tr>
<tr>
<td>TAP (divided attention): delayed reactions</td>
<td>.012</td>
<td>0.70</td>
</tr>
<tr>
<td>WCST: perseverative errors</td>
<td>&lt;.001</td>
<td>0.62</td>
</tr>
<tr>
<td>WCST: perseverative responses</td>
<td>&lt;.001</td>
<td>0.59</td>
</tr>
<tr>
<td>TAP: (focused attention) phasic alertness</td>
<td>.002</td>
<td>0.37</td>
</tr>
<tr>
<td>ANT: Errors of omission</td>
<td>.004</td>
<td>0.31</td>
</tr>
<tr>
<td>WCST: Conceptual level responses</td>
<td>.03</td>
<td>0.31</td>
</tr>
<tr>
<td>CPT-IP: log $\beta$</td>
<td>.004</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Note: DST, Digit Symbol Test; CPT-IP, Continuous Performance Test-Identical Pairs; TAP, Testbatterie zur Aufmerksamkeitsprüfung; IQ, intelligence quotient; ANT, Attention Network Test; WCST, Wisconsin Card Sorting Test.

$^a$Between-group significances as obtained in Kolmogorov-Smirnov test (uncorrected).

$^b$Cohen’s $d$ was computed as $d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$.

Fig. 1. Strong features found in training set. The x axis of each panel is a nonlinear transformation of the basic value shown as $2/\pi \arctan(x/2)$ that is chosen to provide a common visual axis for features that occur on different scales and ranges. The curves and the intensity levels show a smoothed cumulative probability density function for each class, along with the distribution of the absolute difference between these two (red line, the Kolmogorov-Smirnov test statistic represents the maximum value). The figure shows that while similar types of features have similar distribution shapes (e.g., omissions and late reaction in the divided attention test) and vice versa, the strongest differences are distributed among various feature types, suggesting that each type provides complementary information about the diagnostic or subject state. cpt05: CPT-IP hits; d3aug: TAP (divided attention) errors of omission; d3mwg: TAP (divided attention) mean reaction time; speed 1: DST; d3slow: TAP (divided attention) late reactions.
Table 3. Classification Accuracy Measured as Prediction and Validation Error Rates

<table>
<thead>
<tr>
<th>Machine</th>
<th>SVM (Radial Basis Function)</th>
<th>SVM (Linear)</th>
<th>SVM (Radial Basis Function)</th>
<th>LDA</th>
<th>LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE</td>
<td>12.38%</td>
<td>18.13%</td>
<td>13.15%</td>
<td>17.23%</td>
<td>18.60%</td>
</tr>
<tr>
<td>VE ± SD</td>
<td>9.30 ± 2.9%</td>
<td>9.30 ± 3.4%</td>
<td>11.63 ± 3.0%</td>
<td>11.63 ± 3.3%</td>
<td>11.63 ± 3.4%</td>
</tr>
<tr>
<td>Feature 2</td>
<td>TAP-D: E&lt;sub&gt;om&lt;/sub&gt;</td>
<td>TAP-D: RT</td>
<td>TAP-D: RT</td>
<td>TAP-D: RT</td>
<td>TAP-D: RT</td>
</tr>
<tr>
<td>Feature 3</td>
<td>TAP-D: late</td>
<td>DST</td>
<td>TAP-D: late</td>
<td>DST</td>
<td>DST</td>
</tr>
<tr>
<td>Feature 4</td>
<td>TAP-D: late</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: PE, prediction error; VE, validation error; SVM, support vector machine (with different kernels); LDA, linear discriminant analysis; CPT-IP, Continuous Performance Test-Identical Pairs; TAP-D: subtest for divided attention; E<sub>om</sub>, errors of omission; late, late reactions; RT, reaction time; DST, Digit Symbol Test.

Tested methods ranked according to VE. The ± error corresponds to the SD of the binomial distribution at the VE and validation set size (approximately corresponding to a 75% CI). For each machine learning classifier, prediction and VE rates along with selected feature sets are given.

of our combination of attention deficits for diagnostic single-subject classification of schizophrenia.

Most classification studies of schizophrenia used elements of structural or functional magnetic resonance imaging as classifiers, most of them also achieving 80–90% classification accuracy.33–36 Other classification studies focused on electrophysiological methods and obtained classification accuracies of around 70–80%.37–40 Neurocognitive measurements, though most easily accessible and frequently used in both research and clinical routine, have not yet been employed in cross-sectional classification studies. A recent study applied neurocognitive pattern classification to predict conversion to psychosis in at-risk mental states and achieved around 90% classification accuracy in a small and thus preliminary sample.41 By using carefully assembled attention tests that each tap into a different domain, we arrived at a classification accuracy comparable to that reported in the aforementioned studies.

Beyond the classification accuracy, the incremental value of the current study is best considered in the light of a recent criticism expressed by Kapur et al.42 They delineated reasons for the inability of biological psychiatry to establish biomarkers and emphasized that ‘significance chasing’ is frequently performed at the expense of meaningful effect sizes. The implementation of study designs is usually geared toward existing, significant data that either replicate previous findings or that are thought to be replicable. The significance of a biological finding, however, does not necessarily mean that the respective effect is clinically important. Our study addresses this criticism by quantifying the diagnostic accuracy of attention deficits in schizophrenia and thus estimating the hierarchical utility of associated measures. When performing conventional Kolmogorov-Smirnov tests, 16 variables extracted from all performed attention domain tests were statistically significant between the 2 diagnostic groups. However, most of those variables, though statistically significant, remain irrelevant for diagnostic purposes, thus nicely illustrating the difference between statistical significance and clinical usefulness of research data.

The identified discriminators of the present study are consistent with previous findings for the most part. A poorer performance in CPT is reported for both schizophrenia patients and their first-degree relatives.43–45 Measures of hit rate have been widely used as measure of CPT performance and showed high test-retest reliability for schizophrenia patients.43,46,47 Concerning DST, Dickinson and colleagues (2007) have found a very large effect size of 1.57 (Hedges g), which is essentially replicated by our study (Cohen’s d = 1.33).48 Divided attention tasks are rarely applied in schizophrenia research; where it is, the number of omission errors is frequently used to characterize patients with schizophrenia or subjects at high risk of developing psychosis.49 However, there is scarcely existing data on the reliability of this measure. Outcome. Here, our study may serve as a starting point in establishing a classificatory hierarchy of attention deficits in schizophrenia that includes errors of omission as an appropriate measure of divided attention.

There are several limitations to the present study. First, the number of samples can be considered as relatively small for machine learning algorithms and may have resulted in imprecise classification accuracy. On the other hand, and from a clinical point of view, the total of 172 subjects represents a large study population compared to the majority of cognitive studies. Another limitation are the significant between-group differences in education and IQ that may influence attention functions of participants. This serious potential confound is, however, inherent to virtually all cognitive studies into schizophrenia, given the debilitating nature of the disorder. To minimize this confound, mean IQ was entered as a feature in our machine learning analysis but did not contribute to final single-subject classification. Finally, the study focused on attention deficits, while other reported relevant cognitive
deficits in schizophrenia, such as verbal memory or working memory, were not primarily considered. Studies including additional neurocognitive tasks will likely result in further improved classification accuracy. In conclusion, the present study has established a hierarchy of attention deficits in schizophrenia and determined neurocognitive test variables that allow an automated discrimination between healthy and schizophrenic subjects. The study further highlights the potential of machine learning algorithms in the identification of biomarkers that might be employed in addition to clinical assessments.

Acknowledgments

We thank all participants who contributed their time and effort to this study. There is no conflict of interest for any of the authors.

References


