Schizophrenia Bulletin vol. 39 no. 6 pp. 1219–1229, 2013
doi:10.1093/schbul/sbs093
Advance Access publication August 27, 2012

Diagnostic Specificity of Neurophysiological Endophenotypes in Schizophrenia and Bipolar Disorder

Jason K. Johannesen*[,1,2], Brian F. O'Donnell[3,4,5], Anantha Shekhar[4,5], John H. McGrew[6], and William P. Hetrick[3,4,5]

[1]Yale University School of Medicine, Department of Psychiatry, New Haven, CT; [2]VA Connecticut Healthcare System, West Haven, CT; [3]Indiana University, Department of Psychological and Brain Sciences, Bloomington, IN; [4]Indiana University School of Medicine, Department of Psychiatry, Indianapolis, IN; [5]Larue D. Carter Memorial Hospital, Indianapolis, IN; [6]Indiana University-Purdue University Indianapolis, Department of Psychology, Indianapolis, IN

*To whom correspondence should be addressed; VA Connecticut Health Care System, Psychology Service 116-B, 950 Campbell Ave, West Haven, CT 06516, US; tel: 203-932-5711 ext. 2224, fax: 203-937-4735; e-mail: Jason.Johannesen@yale.edu

Background: **The utility of an endophenotype depends on its ability to reduce complex disorders into stable, genetically linked phenotypes. P50 and P300 event-related potential (ERP) measures are endophenotype candidates for schizophrenia; however, their abnormalities are broadly observed across neuropsychiatric disorders. This study examined the diagnostic efficiency of P50 and P300 in schizophrenia as compared with healthy and bipolar disorder samples. Supplemental ERP measures and a multivariate classification approach were evaluated as methods to improve specificity.** ***Methods:*** **Diagnostic classification was first modeled in schizophrenia (SZ = 50) and healthy normal (HN = 50) samples using hierarchical logistic regression with predictors blocked by 4 levels of analysis: (1) P50 suppression, P300 amplitude, and P300 latency; (2) N100 amplitude; (3) evoked spectral power; and (4) P50 and P300 hemispheric asymmetry. The optimal model was cross-validated in a holdout sample (SZ = 34, HN = 31) and tested against a bipolar (BP = 50) sample.** ***Results:*** **P50 and P300 endophenotypes classified SZ from HN with 71% accuracy (sensitivity = .70, specificity = .72) but did not differentiate SZ from BP above chance level. N100 and spectral power measures improved classification accuracy of SZ vs HN to 79% (sensitivity = .78, specificity = .80) and SZ vs BP to 72% (sensitivity = .74, specificity = .70). Cross validation analyses supported the stability of these models.** ***Conclusions:*** **Although traditional P50 and P300 measures failed to differentiate schizophrenia from bipolar participants, N100 and evoked spectral power measures added unique variance to classification models and improved accuracy to nearly the same level achieved in comparison of schizophrenia to healthy individuals.**

Key words: event-related potential/P50/P300/N100/gamma frequency

Introduction

The endophenotype concept posits a characterization of psychiatric disorders informed by traits intermediate to clinical symptoms and underlying, genetically based, pathogenesis.[1] Several neurophysiological measures are currently under investigation as candidate endophenotypes for schizophrenia, of which auditory P50 suppression and P300 event-related potential (ERP) measures have been most extensively studied.[2] However, while in concept these measures may provide a more direct expression of the genetic and pathophysiological underpinnings of neuropsychiatric disorders than symptoms, evidence argues against their specificity to schizophrenia. Recent reviews speak to the overlap between schizophrenia and bipolar disorder on these and other endophenotype candidates,[3,4] raising the question of whether the current diagnostic nosology appropriately characterizes these disorders as distinct etiological or pathophysiological entities.[5] Alternatively, the appearance of overlap between schizophrenia and bipolar disorder may suggest that these measures are sensitive to generalized brain dysfunction but have low specificity as pathognomonic signs of illnesses.

P50 suppression is assessed using a "paired-click" paradigm, in which two identical stimuli are separated by a brief interstimulus interval. P50 suppression refers to the reduction in P50 amplitude to the second stimuli (S2) relative to the first (S1) and is considered an index of the strength of "sensory gating."[6] P50 suppression has been linked to chromosome loci, neurotransmitter systems, and anatomical structures implicated in the neuropathology of schizophrenia.[7,8] P50 suppression deficits have been reported in at-risk individuals[9] and unaffected first-degree relatives of schizophrenia probands,[10] though heritability is evidenced to be lower than other

endophenotype candidates.[11] P50 suppression deficits have also been reported in posttraumatic stress disorder,[12] Huntington's and Parkinson's disease,[13] cocaine dependence,[14] and bipolar disorder[15] among other neuropsychiatric conditions. Taken together, these findings suggest P50 suppression deficits may reflect a vulnerability to, or sequelae of, neuropathology that is common to psychiatric illness broadly.

The auditory P300 ERP is a late positive deflection elicited by infrequent "target" stimuli presented intermittently among more frequently occurring "standard" stimuli, and is generally thought to index the allocation of attention required for stimulus classification. Reductions in P300 peak amplitude and prolongation in peak latency are observed in schizophrenia[16] and in first-degree relatives and individuals with schizotypal features.[17] These abnormalities are reliably detected across schizophrenia subtypes[18] and appear sensitive to illness chronicity.[19] However, as posited for P50 suppression, abnormalities in P300 topography and latency do not appear to reflect specific cerebral disturbances and may occur in any disease that affects cognitive function.[20]

Taken together, it is evident that experimental paradigms used to study P50 and P300 ERPs engage brain circuitry that is disrupted by neuropsychiatric illness. However, there are several reasons to suggest that diagnostic groups compared in these data are inadequately characterized by P50 and P300 responses alone. First, the application of frequency domain analysis to data acquired using the paired-click paradigm has demonstrated superiority to conventional P50 suppression scores when compared on diagnostic discrimination,[21] heritability estimates,[22] and sensitivity to clinical features in schizophrenia.[23,24] Two broad band components representing low-frequency (1–20 Hz) and gamma-band (20–50 Hz) activities have been investigated in studies of P50 suppression following the rationale that mid-latency ERPs reflect the superimposition of these functionally discrete components, indexing early encoding and sensory registration processes, respectively, and that either component can account for differences in ERP amplitude.[21,23,24] Second, using conventional auditory-oddball task data, other studies have identified measures that discriminate schizophrenia from bipolar disorder despite common expression of abnormalities in the posterior–central P300. Specifically, individuals with schizophrenia exhibit amplitude reductions of early (N100, P200, and N200) auditory ERP components[25] and abnormal patterns of lower left than right P300 hemispheric asymmetry[26] that are not observed in bipolar disorder. Finally, beyond the level of individual measures, it has been argued that "characteristic profiles" of multiple indices could be used to enhance diagnostic specificity of ERP endophenotypes.[27] In a recent example of this approach, Price and colleagues[28] constructed a multivariate classification model based on P50, P300, mismatch negativity, and antisaccade errors. This model achieved 78% accuracy classifying schizophrenia against healthy participants, providing a substantial improvement in specificity over any individual endophenotype.

This study evaluated the hypothesis that a multivariate method, combined with thorough characterization of auditory-evoked response data, would improve the diagnostic efficiency of schizophrenia endophenotype candidates. Hierarchical models were first used to evaluate classification accuracy based on candidate endophenotypes (P50 suppression, P300 amplitude, and latency) and then the contributions of supplemental ERP measures (evoked spectral power, N100 amplitude, and hemispheric asymmetry) ascertained from the same experimental paradigms. Using this approach, we aimed to determine (1) the classification accuracy of standard P50 and P300 measures in differentiating schizophrenia from healthy normal adults, (2) whether supplemental ERP measures improve classification accuracy beyond these candidate measures, and (3) the diagnostic specificity of an optimized model for schizophrenia classification against a bipolar disorder comparison group. Although P50 and P300 were expected to classify schizophrenia from healthy normal participants with high accuracy, the supplemental ERP measures were hypothesized to provide superior classification accuracy in differentiating schizophrenia and bipolar disorder.

Method

Participants

A total of 223 cases (schizophrenia, SZ = 88; bipolar disorder, BP = 52; healthy normal, HN = 83) provided written informed consent and completed ERP test procedures in one session. Smaller groups of this sample have been reported in previous studies of P50 suppression.[23,24,29] Participants were 18- to 65-year olds with normal hearing acuity and normal, or corrected-to-normal, visual acuity. Inclusion required a DSM-IV[30] diagnosis of schizophrenia or bipolar disorder I based on structured clinical interview for DSM (SCID)[31] interview and chart review. Healthy normal community volunteers were screened by SCID interview and reported no lifetime Axis I or Axis II psychiatric illness. Excluded were individuals with schizoaffective disorder or major depression without history of mania, current substance dependence, cardiovascular disease, neurological disease, or history of traumatic brain injury.

The distributional properties of all ERP data were examined for normality following square root transformation where appropriate. The Boxplot function of SPSS (version 14.0.1, SPSS Inc., 2005) was used to identify extreme outliers according to the standard criteria of 3

box lengths beyond the interquartile range. In total, 4 SZ, 2 BP, and 2 HN participants were excluded from the sample due to outlying values on one or more ERP measure. To obtain equal prior probability of group membership in classification analyses and to reduce the influence of age on ERP measures, SZ and HN participants were age matched 1:1 to the 50 retained BP participants. Gender was used as a secondary matching criterion where possible. Descriptive statistics of the final matched samples are presented in table 1. The remaining 34 SZ and 31 HN participants were used as a holdout sample for cross-validation of classification models based on the matched samples.

Symptom Measures

Symptom severity was assessed in the SZ sample using the Positive and Negative Syndrome Scale,[32] scored according to the 5-factor structure of Bell and colleagues.[33] Symptom severity was assessed in the BP sample using the Young Mania Rating Scale[34] and Montgomery Asberg Depression Rating Scale.[35]

ERP Procedures

Auditory-evoked responses were ascertained using 2 experimental tasks, a paired-click (P50 suppression) and a 2-stimulus auditory discrimination (P300) task. The paired-click procedure consisted of 130 binaural-paired auditory click trials (7–11 s inter-trial interval (ITI), 500 ms inter-stimulus interval) presented through insert earphones (peak intensity = 81 dB SPL; 3 ms; 58 dB SPL white-noise background). Participants were seated upright and responded by key press to infrequent paired tones ($N = 20$), easily distinguished from test stimuli ($N = 110$) on the basis of tone frequency. These "infrequent" trials were included to maintain alertness and a consistent level of engagement in the task across participants[23,24] but were not scored for statistical analysis. The auditory discrimination task consisted of 75 targets (1500 Hz tone) randomly interspersed among 425 standard (1000 Hz tone) trials. Both stimuli were presented at an intensity of 86 dB SPL and a duration of 50 ms, separated by a 1.2-s ITI. Participants responded to targets by button press, with response randomly assigned to the right or left index finger.

Table 1. Sample Descriptive Statistics

Variable	HN, $n = 50$	BP, $n = 50$	SZ, $n = 50$	Comparison
Age: *M* (SD)	38.6 (9.6)	40.7 (11.2)	40.5 (10.1)	SZ = HN; SZ = BP
Gender: % male	50%	44%	64%	SZ = HN; SZ ≠ BP*
Handedness: % R	92%	90%	92%	SZ = HN; SZ = BP
Substance use past 24 h: % used				
Nicotine	33%[a]	48%	49%[a]	SZ ≠ HN*; SZ = BP
Alcohol	4%[a]	0%	6%[a]	SZ = HN; SZ = BP
Drugs	0%[a]	2%	4%[a]	SZ = HN; SZ = BP
PANSS: *M* (SD)[b]				
Positive			16.2 (5.8)	
Negative			15.9 (6.4)	
Cognitive			15.7 (4.4)	
Hostility			5.6 (2.1)	
Emotional			8.9 (4.3)	
Total			64.3 (15.9)	
MADRS: *M* (SD)[c]		14.8 (9.1)		
YMRS: *M* (SD)[d]		18.1 (9.7)		
Medication: % RX				
Atypical[e]	0%	34%	66%	SZ ≠ BP**
Traditional[e]	0%	4%	18%	SZ ≠ BP*
Anticholinergic	0%	8%	16%	SZ = BP
Antidepressant	0%	16%	26%	SZ = BP
Anticonvulsant	0%	42%	20%	SZ ≠ BP*
Anxiolytic	0%	40%	16%	SZ ≠ BP**
Withdrawn	NA	36%	18%	SZ ≠ BP*

Note: HN, healthy normal; BP, bipolar; SZ, schizophrenia, R, right handed; RX, treatment prescribed; PANSS, Positive and Negative Syndrome Scale; MADRS, Montgomery-Asberg Depression Rating Scale; YMRS, Young Mania Rating Scale; NA, not applicable.
[a]Data available for 49 participants.
[b]Data available for 36 participants.
[c]Data available for 44 participants.
[d]Data available for 47 participants.
[e]Class of antipsychotic medication.
*$P < .05$, **$P < .01$ (two-tailed).

Electroencephalographic (EEG) data were recorded with a 32-channel cap (10–20 system; Falk Minow Services, Munich, Germany) and bioamplification system (SynAmps, Neuroscan Inc., Sterling, VA). Vertical (VEOG) and horizontal (HEOG) eye movements were recorded (1000 gain) for offline ocular correction. For the P50, data were acquired at a 1000-Hz sampling rate with an analog high-pass filter of .10 Hz and low-pass filter of 200 Hz. For the P300, the sampling rate and low-pass settings were the same, but the high-pass filter was set to .05 Hz. In both procedures, gain was 5000 and cortical leads were referenced to the nose. Impedances were maintained below 10 000 Ohms during recording. P50 and P300 procedures were administered in pseudorandom order.

Data analysis was conducted using Brain Vision Analyzer software (Brain Products, Munich, Germany). P50 data were processed by segmenting continuous EEG into 450-ms epochs beginning 100 ms before stimulus onset, baseline correcting, and bandpass filtering from 1 to50 Hz (48 dB/octave) prior to ocular artifact correction.[36] Epochs containing activity with a voltage range of 150 µV at electrode FCz were excluded, and data were manually inspected for residual artifact and flat-lined trials before averaging. The P50 ERP was identified at FCz as the largest positive deflection in the average waveform from 40 to 80 ms. P50 peak amplitude was measured relative to a preceding trough 35–50 ms post stimulus. Semiautomated peak detection was applied, with markers set automatically according to latency criteria and visually inspected before accepted for analysis. P50 peak amplitude was scored at bilateral left (T7) and right (T8) temporal electrodes using the same criteria. P50 suppression was computed using the S1 − S2 difference score, found to be superior to the S2/S1 ratio score in terms of psychometric stability[37] and heritability estimates.[11] Hemispheric asymmetry of P50 amplitude to S1 was computed using a difference score [T8–T7], which simplifies analysis of asymmetry to a single variable and reduces the influence of differences in overall signal strength between groups.[38] N100 was measured from S1 trials of the paired-click paradigm after applying a 20-Hz (48 dB/octave roll-off) low-pass filter to the average waveform. The N100 peak was identified at electrode Cz within a 80- to 150-ms latency window, and peak amplitude was measured relative to prestimulus baseline. Finally, following initial processing, S1 trials were resegmented from 0 to 256 ms, filtered (48 dB/octave) in 1- to 20-Hz bins and 20- to 50-Hz bins, and averaged. Averaged data were submitted to Fast Fourier Transform with Hamming windowing (10%) to extract evoked spectral power of low-frequency (LFR, 1–20 Hz) and gamma-band (GBR, 20–50 Hz) responses as described in previous work.[21,23,24] Electrode FCz was selected for analysis of the GBR, as a frequency domain analog to the P50 ERP, and electrode Cz was selected for the LFR, as an analog to the N100 ERP.

The P300 ERP was scored by segmenting EEG into 900 ms epochs with a 100-ms prestimulus baseline for standard and target trials separately. The same procedures used for artifact rejection and blink correction of P50 data were applied to these data. Continuous data were low-pass filtered at 20 Hz (48 dB/octave) and averaged for peak detection. P300 peak amplitude was measured relative to a 100 ms baseline as the largest positive deflection 280–600 ms poststimulus at electrode Pz. Peak detection for analysis of hemispheric asymmetry and N100 used the same methods and electrode sites as described for paired-click paradigm data. Finally, the GBR and LFR were extracted separately from standard and target tone trials using the same procedures applied to paired-click paradigm data.

Statistical Analysis

Two data analytic strategies were applied, following the method of Price and colleagues.[28] First, age-matched SZ and HN groups were compared on all ERP measures using independent samples *t*-tests. Because these analyses were used to replicate commonly reported findings in literature and the direction of expected effects was known a priori, one-tailed tests were used without correction for multiple comparisons. The effect size of each contrast was computed using Cohen's *d* statistic.[39] The frequency distribution of deficit on each ERP measure was then examined in SZ and BP samples, with deficit defined as 1 standard deviation (SD) below the mean (16th percentile) of the HN score distribution. Proportional differences in deficit status between SZ and BP samples were tested nonparametrically using Chi square.

Following a second analytic approach, diagnostic classification models were constructed by entering ERP variables into multivariate logistic regressions (cut .5) as predictors of group membership. Predictors were clustered in blocks ordered[1] by level of analysis: Block 1 = P50 suppression, P300 amplitude, and P300 latency; Block 2 = N100 amplitude (3 measures); Block 3 = evoked spectral power (GBR: 3 measures; LFR: 3 measures); and Block 4 = hemispheric asymmetry (2 measures). To accommodate the large number of independent variables (14 total) relative to sample size (N = 100 per comparison), conditional entry of predictors was used at each block. Using backward entry, the first step of the analysis tested the independent and combined contributions of the 3 candidate endophenotypes, with entry of predictors in subsequent blocks conditional on improvement in likelihood estimates over preceding blocks. This procedure was first used to classify SZ against the HN sample. The optimal classification model was cross-validated in the matched sample by discriminant function analysis using the leave-one-out method, and again by applying the discriminant function to the holdout sample. Diagnostic

specificity of the optimal multivariate model was then tested in classification analyses of SZ against BP.

Results

Sample Characteristics and Descriptive Statistics

SZ and HN differed in the proportion of participants reporting nicotine use in the 24-h period prior to EEG data collection. Differences in gender distribution and in medication types were observed between SZ and BP. Accordingly, nicotine use, gender, and medication type were tested as covariates of ERP predictor variables in classification analyses where appropriate. Group descriptive statistics are presented in table 1.

Univariate Contrasts: SZ vs HN

Statistical results of paired contrasts described below, including effect size estimates, are presented in table 2. ERP Waveforms are presented in online supplementary material figure A.

ERP Peak Amplitude. SZ evidenced a trend-level difference [$t(98) = 1.59$, $P = .057$] in P50 suppression compared with HN.[2] This effect appeared to be dependent on P50 amplitude in response to S1, which correlated highly with computed suppression values across the sample [$r(150) = .858$, $P < .001$] and was smaller in SZ, though only at trend level [$t(98) = 1.40$, $P = .081$].

Large effect size differences were obtained in comparison of N100 responses to S1 in the paired-click paradigm and to both standard and target tones from the oddball paradigm.

Medium effect size differences were obtained for P300 peak amplitude and latency measures. Both differed significantly between groups in the expected directions, with smaller P300 amplitude and later peak latency in SZ.

Small but significant effects were observed for P50 and P300 hemispheric asymmetry, characterized by larger right than left hemisphere amplitude in SZ and larger left than right hemisphere amplitude in HN. Differences in asymmetry were further tested by repeated-measures analysis of variation, yielding a significant group × hemisphere interaction [$F(1,98) = 5.14$, $P = .024$] for P50 but only a trend-level effect for P300 [$F(1,98) = 3.39$, $P = .07$].

Evoked Spectral Power. SZ evidenced significant reductions in both GBR and LFR spectral power measures in

Table 2. Comparison of Schizophrenia and Healthy Normal Group

Measure	HN, $n = 50$ (M, SD)	SZ, $n = 50$ (M, SD)	$t(98)$[a]	Effect (d)[b]
Paired-click Paradigm				
Amplitude				
P50 S1	4.96 (3.00)	4.14 (2.82)	1.40	.28 (S)
P50 S2	2.06 (1.66)	1.97 (1.65)	.27	.05 (S)
P50 suppression (S1 − S2)	2.90 (2.45)	2.18 (2.11)	1.59	.32 (S)
N100 S1	−11.80 (6.38)	−6.60 (5.03)	−4.52***	.91 (L)
Spectral power[c]				
GBR S1	.31 (.15)	.25 (.11)	2.31*	.47 (S)
LFR S1	3.11 (1.41)	1.84 (1.00)	5.18***	1.05 (L)
Hemispheric asymmetry				
P50 asymmetry	−.41 (2.02)	.32 (1.08)	−2.27*	.46 (S)
Auditory Oddball Paradigm				
Amplitude				
P300 amplitude	15.07 (7.17)	10.74 (6.01)	3.28***	.66 (M)
P300 latency[c]	19.06 (1.11)	20.19 (1.77)	−3.84***	.77 (M)
N100 standard	−3.05 (1.78)	−1.68 (1.18)	−4.53***	.91 (L)
N100 target	−4.42 (2.81)	−2.17 (1.88)	−4.71***	.95 (L)
Spectral power[c]				
GBR standard	.07 (.03)	.05 (.02)	3.03***	.61 (M)
GBR target	.13 (.06)	.10 (.04)	2.89***	.58 (M)
LFR standard	.73 (.33)	.51 (.29)	3.48***	.70 (M)
LFR target	1.25 (.57)	.75 (.30)	5.42***	1.09 (L)
Hemispheric asymmetry				
P300 asymmetry	−.58 (2.25)	.17 (1.82)	−1.84*	.37 (S)

[a]Independent samples t-tests are one-tailed.
[b]Effect size estimates were computed using Cohen's d statistic: $M_1 - M_2/\sigma_{pooled}$ (Cohen, 1988). Positive effect sizes indicate smaller ERP values in SZ relative to HN. Effect size interpretation is based on the following guidelines: <.50 = small (S), .50–.80 = medium (M), >.81 = large (L).
[c]Analysis based on square root-transformed values.
*$P < .05$, ***$P < .001$.

Table 3. Deficit Distribution

	P50-Paired Click Variables						P300-Oddball Variables								
	P50 S1	P50 S1 − S2	N100 S1	GBR S1[a]	LFR S1[a]	P50 Asymmetry	P300 Amp	P300 Latency[a]	N100 Std	N100 Target	GBR Std[a]	GBR Target[a]	LFR Std[a]	LFR Target[a]	P300 Asymmetry
Cut point	Below	Below	Above	Below	Below	Above	Below	Above	Above	Above	Below	Below	Below	Below	Above
	1.96	.45	−5.42	.16	1.70	1.61	7.90	20.17	−1.27	−1.61	.04	.07	.40	.68	1.67
BP deficit	20%	16%	28%	14%	34%	16%	32%	32%	18%	24%	12%	6%	26%	26%	14%
SZ deficit	28%	20%	44%	20%	54%	10%	34%	38%	42%	38%	28%	18%	40%	42%	16%
χ^2	.88	.27	2.78	.64	4.06*	.80	.04	.40	6.86**	2.29	4.00*	3.41	2.22	2.85	.08

Note: χ^2 tests are two-tailed, df = 1.
[a]Analysis based on square root-transformed values.
*$P < .05$, **$P < .01$.

response to S1 trials of the paired-click paradigm and to both standard and target trials of the auditory oddball paradigm. All tests were significant ($P < .05$) but ranged in effect size from small for GBR to S1 to large for LFR to oddball targets.

Deficit Distribution. Proportions of SZ and BP participants classified as deficit by individual measures were generally below 50% and comparable between groups with 3 exceptions (table 3). SZ evidenced proportionately higher rates of deficit on LFR to S1 in the paired-click paradigm and on N100 and GBR spectral power to standard tone trials of the auditory oddball paradigm.

Interdependence of ERP Predictors

Interrelationships between ERP predictor measures (see online supplementary material table B) were examined for collinearity prior to classification analyses. Two variables were identified for further inspection using collinearity diagnostics based on substantial shared variance, N100 to S1 paired-click trials and the LFR to the same stimuli ($r = -.863$). Tolerance and variance inflation factor (VIF) were examined by regressing the remaining ERP measures on these 2 variables in separate models. In both models, tolerance was above .4 and VIF below 2.3 across all variables, suggesting multicollinearity was not significant according to conventional guidelines of tolerance <.25 or VIF >4.[40]

Multivariate Classification Analyses

Schizophrenia vs Healthy. P50 suppression,[3] P300 amplitude, and P300 latency together provided 71% overall accuracy classifying SZ against HN (table 4). P50 suppression failed to contribute significantly and was removed, reducing overall classification accuracy to 70%. In subsequent blocks, N100 to target trials, LFR to target trials, and GBR to standard trials in the oddball procedure entered the model (table 5) and increased classification accuracy to 79% (78% for SZ, 80% for HN). Hemispheric asymmetry failed to enter the model, entered as T8 − T7 difference score or when retested as the interaction of T8 × T7. The final model provided likelihood ratios of 3.9 and .3 for positive (predicting SZ) and negative test results, respectively, compared to ratios of 2.5 and .4 obtained by the initial model.

A significant relationship between nicotine use in past 24 h and P300 amplitude was detected [$r(98) = -.24$, $P = .02$]. Entering nicotine use as a predictor did not alter final ERP model parameters.

Two cross-validation methods were applied to evaluate the stability of the final classification model. First, the 5 ERP predictor variables retained in this model were combined in a single discriminant function classifying the matched SZ and HN subjects ($n = 100$). Overall

Table 4. Classification Analysis of SZ vs HN

Predictors		B	SE	Wald	df	Sig.	Exp(B)	Diagnostic Efficiency		
								Sensitivity	Specificity	Accuracy
Candidate endophenotypes								70%	72%	71%
Block 1	P300 latency	.542	.184	8.713	1	.003	1.720			
	P300 Amp	−.091	.036	6.513	1	.011	.913			
	P50 S1 − S2	−.052	.102	.256	1	.613	.950			
	Constant	−9.293	3.612	6.619	1	.010	.000			
Optimal model								78%	80%	79%
Block 1	P300 latency	.415	.201	4.263	1	.039	1.514			
	P300 Amp	−.072	.043	2.744	1	.098	.931			
Block 2	N100 target	.320	.144	4.929	1	.026	1.378			
Block 3	LFR target	−2.060	.842	5.988	1	.014	.127			
	GBR standard	−20.752	11.547	3.230	1	.072	.000			
	Constant	−3.003	3.967	.573	1	.449	.050			

Table 5. Classification Analysis of SZ vs BP

Predictors		*B*	SE	Wald	df	Sig.	Exp(B)	Diagnostic Efficiency		
								Sensitivity	Specificity	Accuracy
Candidate endophenotypes								54%	44%	49%
Block 1	P300 latency	.046	.128	.130	1	.718	1.047			
	P300 Amp	−.024	.034	.491	1	.483	.977			
	P50 S1 − S2	−.040	.86	.216	1	.642	.961			
SZ vs HN optimal model								70%	58%	64%
Block 2	N100 target	.231	.107	4.689	1	.030	1.259			
Block 3	GBR standard	−22.631	9.807	5.325	1	.021	.000			
	Constant	1.954	.673	8.422	1	.004	7.055			
SZ vs BP optimal model								74%	70%	72%
Block 2	N100 S1	.099	.043	5.401	1	.020	1.104			
	N100 Target	.383	.187	4.191	1	.041	1.467			
Block 3	GBR Target	−14.940	5.017	8.869	1	.003	.000			
	Constant	3.387	.819	17.121	1	.000	29.575			

Note: Order of entry for Blocks 2–4 was guided by effect size of contrasts comparing SZ and HN across level of analysis, with better discriminating tests entered first.
P50 suppression was also tested using the common S2/S1 ratio method: $t(98) = -.56$, $P = .29$ (one-tailed).
Classification analysis was repeated entering P50 S1 amplitude in place of P50 suppression (S1 − S2). As with P50 suppression, P50 S1 amplitude failed to enter the model significantly [Wald(1) = .121, $P = .73$], and was dropped after the first step.

classification accuracy of 76% was achieved, with 75% accuracy attained using leave-one-out cross-validation. Second, this analysis was repeated using a holdout sample of 34 SZ and 31 HN participants. Classification accuracy was nearly identical to the original sample, with 77% and 74% accuracy for the full sample and leave-one-out validation, respectively.

Schizophrenia vs Bipolar. An initial classification analysis based on candidate endophenotype measures failed to differentiate SZ from BP above chance level (table 5). The multivariate model derived from classification of SZ against HN provided 64% overall accuracy when used to classify SZ (70%) against BP (58%). However, of the 5 predictors tested, only N100 to target trials and GBR to standard trials were retained.

BP and SZ groups were not evenly matched by gender distribution. A significant correlation between gender and target trial N100 (larger in women) was detected [$r(100) = -.26$, $P = .008$]; however, gender did not enter significantly into the regression model or alter coefficients for the 2 ERP predictors. BP and SZ participants also differed in proportions prescribed atypical and traditional antipsychotic, anticonvulsant, and anxiolytic medications. Medication status entered into the regression and increased classification accuracy to 78% overall, with 78% of BP and 78% of SZ assigned correctly. This result is not scientifically meaningful, because differences in medication

type are considered conditional on diagnosis. Importantly, the original parameters were retained in this model, thus appearing to be stable predictors even when accounting for differences in medication between patient groups.

To determine whether higher classification could be achieved using an alternative set of predictors, an unrestricted model including all predictors was tested in classification of SZ and BP. The optimal model achieved 72% accuracy overall (74% for SZ, 70% for BP) based on 3 predictors: N100 to S1 trials of the paired-click paradigm, N100 to standard trials of the auditory oddball paradigm, and GBR to target trials of the auditory oddball paradigm (table 5). This optimal model provided likelihood ratios of 2.5 and .4 for positive (predicting SZ) and negative test results, respectively, compared to ratios of 1 and 1 obtained by the initial model based on P50 and P300 measures alone. This model was again tested by discriminant function analysis, producing 69% overall classification accuracy for the full sample and again when tested using leave-one-out validation.

Discussion

This study evaluated the diagnostic efficiency of P50 and P300 endophenotype candidates for schizophrenia. When entered together in classification analyses comparing schizophrenia against the healthy sample, correct group membership was predicted with 71% accuracy overall, with removal of P50 suppression only decreasing classification accuracy to 70%. This level of accuracy is comparable to that achieved by Price and colleagues[28] using P300 amplitude alone (72%), but lower than that attained by a multivariate endophenotype (overall accuracy = 78%) based on 4 endophenotype measures. Similar classification accuracy (79%) was achieved by the optimal multivariate model constructed in this study, based on P300 amplitude and latency and 3 supplemental auditory-oddball task measures (N100 to targets, LFR to targets, and GBR to standards). Accordingly, results support the premise that diagnostic efficiency is enhanced by multivariate information. On a practical note, this improvement was achieved in ERP data collected in a standard 2-stimulus auditory discrimination task and thus did not require additional psychophysiological data collection.

A central issue guiding this analysis concerns the extent of overlap between schizophrenia and bipolar disorder on P50 and P300 endophenotypes, and the possibility that diagnostic distinctions could be enhanced with more thorough characterization of these ERP measures. Given evidence that individuals with bipolar disorder exhibit deficits similar to schizophrenia on P50 and P300 measures,[15,25] no appreciable separation between these patient samples was expected. Indeed, nearly equivalent proportions of schizophrenia and bipolar participants exhibited a P50 suppression deficit of at least 1 SD below the healthy normal average. The frequency of deficit was higher, but proportionately comparable between groups, when assessed by P300 amplitude and latency. Based on these profiles, the candidate endophenotypes appeared to be equally sensitive to deficits in schizophrenia and bipolar disorder, but specific to neither. This observation was further supported in classification analysis based only on P50 suppression and P300 amplitude and latency, which failed to differentiate the patient samples beyond chance level.

How do these findings align with arguments that schizophrenia and bipolar disorder share common pathophysiological deficits? Overall, distributions of deficit in schizophrenia and bipolar samples were comparable with schizophrenia, evidencing significantly higher rates of deficit on only 3 measures: LFR to S1 in the paired-click paradigm, N100 to standard tones in the oddball paradigm, and GBR to standard tones in the oddball paradigm. When the multivariate model developed in reference to the healthy normal group was applied to the patient samples, 2 of the predictors (N100 to target trials and GBR to standard trials) were retained and distinguished the patient samples with 64% classification accuracy. Classification accuracy was further improved to 72% by a model based specifically on multivariate differences in schizophrenia and bipolar sample data, including N100 from the paired-click paradigm, N100 to standards, and the GBR to target trials. Taken together, the appearance of common pathophysiological deficit in schizophrenia and bipolar disorder appears dependent on the measure by which this comparison is made. Although our results are in agreement with prior work postulating that P50 and P300 are similarly impaired, our conclusions based on more thorough characterization of psychophysiological data favor a diagnostic distinction that is only slightly less robust than that between schizophrenia and healthy normal adults.

Several secondary findings warrant discussion. First, P50 suppression provided little appreciable contribution to classification analyses. Despite encouraging results from early studies of P50 suppression in schizophrenia, more recent evidence suggests that heterogeneity in illness expression[14,24] and methodological differences in the ascertainment of P50 between laboratories[41,42] influence the magnitude of differences observed between schizophrenia and healthy comparison subjects. There is also contrary evidence regarding the trait stability of P50 suppression deficits, including reports of normal P50 suppression in first-episode schizophrenia patients and their unaffected siblings,[43] in medication-withdrawn patients,[44] and in chronic patients treated with atypical antipsychotic medications.[45] Second, it is notable that in contrast to P50 suppression findings, N100 amplitude to S1, as well as to standard and target oddball paradigm trials, robustly and consistently differed between schizophrenia and healthy normal subjects. Variability in findings of N100 deficit in

schizophrenia has previously been discussed in relation to experimental (ie, interstimulus interval) and subject factors (ie, medication status).[46] While the auditory stimuli (rarefaction click vs 1000 Hz tone) and interstimulus intervals (8- to 11-s ITI vs 1.2-s ITI) used in paired-click and oddball paradigms evidently influenced the amplitude of N100 amplitude between experiments, the magnitude of effects distinguishing schizophrenia from healthy groups was remarkably similar across N100 measures. Therefore, differences in stimuli and task parameters did not influence the detection of N100 amplitude reductions in schizophrenia or effect sizes when compared with healthy participants. Furthermore, classification models distinguishing schizophrenia from bipolar disorder were unaffected by the inclusion of medication status as a covariate. Third, the GBR and LFR provided additional measures that were correlated with, but not redundant to, conventional ERP amplitudes. Correlation coefficients indicate high association between P50 and the GBR in the paired-click paradigm, and between N100 and the LFR when ascertained in response to common stimuli. However, effect size contrasts suggest that the evoked frequency domain measures generally provided stronger group separation between schizophrenia and healthy samples than their associated ERP amplitude measures. Therefore, although evoked power measures do share variance with conventional ERP measures, the frequency analytic approach appeared to be a more sensitive measure in these comparisons. The advantage of the frequency analytic approach may be attributed to more precise characterization of evoked activity as functionally distinct neural ensembles,[21] which are superimposed in ERP averages, as well as improvements in reliability that could be expected in measures of the magnitude of a waveform across a broad sample of time (ie, 256-ms epoch) vs at a single peak. A final ERP measure, hemispheric asymmetry, was used to examine differences in the spatial distribution of ERP generation at right and left temporal scalp regions. Although asymmetry values did not enter classification models, observed effects were nonetheless consistent with previous reports of left hemisphere reductions in P50[47] and auditory P300[26] in schizophrenia.

Several limitations warrant consideration in interpreting these results. First, it does appear that although not assessed in these data, groups may differ in P200 and N200 ERPs (see online supplementary material figure A). Second, frequency domain analysis based on broadly defined LFR (1–20 Hz) and GBR (20–50 Hz) components may obscure important differences that exist within traditional delta, theta, alpha, beta, and gamma ranges. Moreover, time-frequency approaches to data analysis provide additional information concerning trial-to-trial consistency in power and phase that could add further precision to our characterization of evoked response data. Taken together, we cannot conclude that the models developed on selected measures are indeed optimal, because other important differences in time (ie, P200 and N200) and frequency domain responses may further enhance diagnostic differences reported in this study. Finally, medication, substance use history, and medical comorbidities should not be overlooked as potential confounds of any study attempting to use neurophysiological measures to classify individuals with severe psychiatric illnesses. We attempted to address these concerns in covariate analysis, but statistical approaches cannot adequately account for the independent and combined effects of these features of illness chronicity on brain development and function.

In summary, classification models based on multiple sources of auditory-evoked response information enhanced diagnostic efficiency over P50 and P300 endophenotype candidates. Whereas the present analysis of P50 and P300 measures failed to distinguish schizophrenia from bipolar participants above chance level, N100 ERP and GBR spectral power measures discriminated schizophrenia from bipolar disorder at nearly as high accuracy as from healthy participants. We interpret these results to suggest that fundamental differences in stimulus registration indexed by the GBR and early attention-perceptual processes associated with N100 distinguish schizophrenia from bipolar disorder, in which case dysfunction is relatively more prominent in schizophrenia. This finding is also important to consider in interpretation of P300 abnormalities that appear common to both disorders. Although P300 is thought to reflect the allocation of cognitive resources to stimulus evaluation through temporal–parietal structures, it is reasonable to speculate that abnormalities observed in schizophrenia are sensory-perceptual in nature and originate along thalamocortical projections to the temporal lobe. Future progress in the investigation of psychophysiological endophenotypes may require more thorough characterization of underlying properties of evoked response data and a departure from single peak amplitude measures. Multiple, non-overlapping components in time and frequency domains should be examined as independent, and perhaps more reliable, indices of information processing abnormalities associated with commonly studied ERP endophenotypes.

Funding

Young Investigator Award from NARSAD (to WPH); National Institute of Mental Health grants (R03 MH066149-01A1, R01 MH074983-01 to WPH and R21MH071876-01 to BFO); and VA Rehabilitation Research & Development (D7008W to JKJ). There are no conflicts of interest for any of the authors of this paper. No author has any possible financial gain for the findings presented here.

Supplementary Material

Supplementary material is available at http://schizophreniabulletin.oxfordjournals.org.

References

1. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry*. 2003;160:636–645.
2. Turetsky BI, Calkins ME, Light GA, Olincy A, Radant AD, Swerdlow NR. Neurophysiological endophenotypes of schizophrenia: the viability of selected candidate measures. *Schizophr Bull*. 2007;33:69–94.
3. Thaker GK. Neurophysiological endophenotypes across bipolar and schizophrenia psychosis. *Schizophr Bull*. 2008;34:760–773.
4. Lin P-I, Mitchell BD. Approaches for unraveling the joint genetic determinants of schizophrenia and bipolar disorder. *Schizophr Bull*. 2008;34:791–797.
5. Maier W, Zobel A, Wagner M. Schizophrenia and bipolar disorder: differences and overlaps. *Curr Opin Psychiatry*. 2006;19:165–170.
6. Freedman R, Adler LE, Waldo MC, Pachtman E, Franks RD. Neurophysiological evidence for a defect in inhibitory pathways in schizophrenia: comparison of medicated and drug-free patients. *Biol Psychiatry*. 1983;18:537–551.
7. Freedman R, Olincy A, Ross RG, et al. The genetics of sensory gating deficits in schizophrenia. *Curr Psychiatry Rep*. 2003;5:155–161.
8. Leonard S, Gault J, Hopkins J, et al. Association of promoter variants in the alpha7 nicotinic acetylcholine receptor subunit gene with an inhibitory deficit found in schizophrenia. *Arch Gen Psychiatry*. 2002;59:1085–1096.
9. Cadenhead KS, Light GA, Geyer MA, Braff DL. Sensory gating deficits assessed by the P50 event-related potential in subjects with schizotypal personality disorder. *Am J Psychiatry*. 2000;157:55–59.
10. Clementz BA, Geyer MA, Braff DL. Poor P50 suppression among schizophrenia patients and their first-degree biological relatives. *Am J Psychiatry*. 1998;155:1691–1694.
11. Greenwood TA, Braff DL, Light GA, et al. Initial heritability analyses of endophenotypic measures for schizophrenia: the consortium on the genetics of schizophrenia. *Arch Gen Psychiatry*. 2007;64:1242–1250.
12. Neylan TC, Fletcher DJ, Lenoci M, et al. Sensory gating in chronic posttraumatic stress disorder: reduced auditory P50 suppression in combat veterans. *Biol Psychiatry*. 1999;46:1656–1664.
13. Skinner RD, Miyazato H, Garcia-Rill E. The sleep state-dependent P50 auditory evoked potential in neuropsychiatric diseases. *Int Congr Ser*. 2002;1232:813–825.
14. Boutros N, Zouridakis G, Rustin T, Peabody C, Warner D. The P50 component of the auditory evoked potential and subtypes of schizophrenia. *Psychiatry Res*. 1993;47:243–254.
15. Schulze KK, Hall M, McDonald C, et al. P50 auditory evoked potential suppression in bipolar disorder patients with psychotic features and their unaffected relatives. *Biol Psychiatry*. 2007;62:121–128.
16. Jeon Y-W, Polich J. Meta-analysis of P300 and schizophrenia: patients, paradigms, and practical implications. *Psychophysiology*. 2003;40:684–701.
17. Kimble M, Lyons M, O'Donnell B, Nestor P, Niznikiewicz M, Toomey R. The effect of family status and schizotypy on electrophysiologic measures of attention and semantic processing. *Biol Psychiatry*. 2000;47:402–412.
18. St Clair D, Blackwood D, Muir W. P300 abnormality in schizophrenic subtypes. *J Psychiatric Res*. 1989;23:49–55.
19. Mathalon DH, Ford JM, Rosenbloom M, Pfefferbaum A. P300 reduction and prolongation with illness duration in schizophrenia. *Biol Psychiatry*. 2000;47:413–427.
20. Hamburger HL, Triantafyllou NI. Clinical applications of auditory event related potentials in neurology. *Brain Topogr*. 1990;3:49–52.
21. Clementz BA, Blumenfeld LD. Multichannel electroencephalographic assessment of auditory evoked response suppression in schizophrenia. *Exp Brain Res*. 2001;139:377–390.
22. Hong LE, Summerfelt A, Mitchell BD, et al. Sensory gating endophenotype based on its neural oscillatory pattern and heritability estimate. *Arch Gen Psychiatry*. 2008;65:1008–1016.
23. Johannesen JK, Bodkins M, O'Donnell BF, Shekhar A, Hetrick WP. Perceptual anomalies in schizophrenia co-occur with selective impairments in the gamma frequency component of midlatency auditory ERPs. *J Abnorm Psychol*. 2008;117:106–118.
24. Johannesen JK, Kieffaber PD, O'Donnell BF, Shekhar A, Evans JD, Hetrick WP. Contributions of subtype and spectral frequency analyses to the study of P50 ERP amplitude and suppression in schizophrenia. *Schizophr Res*. 2005;78:269–284.
25. O'Donnell BF, Vohs JL, Hetrick WP, Carroll CA, Shekhar A. Auditory event-related potential abnormalities in bipolar disorder and schizophrenia. *Int J Psychophysiol*. 2004;53:45–55.
26. Salisbury DF, Shenton ME, McCarley RW. P300 topography differs in schizophrenia and manic psychosis. *Biol Psychiatry*. 1999;45:98–106.
27. Friedman D, Squires-Wheeler E. Event-related potentials (ERPs) as indicators of risk for schizophrenia. *Schizophr Bull*. 1994;20:63–74.
28. Price GW, Michie PT, Johnston J, et al. A multivariate electrophysiological endophenotype, from a unitary cohort, shows greater research utility than any single feature in the Western Australian family study of schizophrenia. *Biol Psychiatry*. 2006;60:1–10.
29. Carroll CA, Kieffaber PD, Vohs JL, O'Donnell BF, Shekhar A, Hetrick WP. Contributions of spectral frequency analyses to the study of P50 ERP amplitude and suppression in bipolar disorder with or without a history of psychosis. *Bipolar Disord*. 2008;10:776–787.
30. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: Fourth Edition*. Washington, DC: American Psychiatric Press; 1994.
31. First M, Spitzer R, Gibbon M, Williams J. *Structured Clinical Interview for DSM-IV Axis I Disorders: Patient Edition.* SCID - I/P. New York: New York State Psychiatric Institute; 1996.
32. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13:261–276.
33. Bell MD, Lysaker PH, Beam-Goulet JL, Milstein RM, Lindenmayer JP. Five-component model of schizophrenia: assessing the factorial invariance of the positive and negative syndrome scale. *Psychiatry Res*. 1994;52:295–303.
34. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry*. 1978;133:429–435.

35. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134:382–389.
36. Gratton G, Coles MG, Donchin E. A new method for off-line removal of ocular artifact. *Electroencephalogr Clin Neurophysiol*. 1983;55:468–484.
37. Smith DA, Boutros NN, Schwarzkopf SB. Reliability of P50 auditory event-related potential indices of sensory gating. *Psychophysiology*. 1994;31:495–502.
38. Allen JJB, Coan JA, Nazarian M. Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion. *Biol Psychol*. 2004;67:183–218.
39. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
40. O'brien R. A caution regarding rules of thumb for variance inflation factors. *Qual Quant*. 2007;41:673–690.
41. Patterson JV, Hetrick WP, Boutros NN, et al. P50 sensory gating ratios in schizophrenics and controls: a review and data analysis. *Psychiatry Res*. 2008;158:226–247.
42. de Wilde OM, Bour LJ, Dingemans PM, Koelman JHTM, Linszen DH. A meta-analysis of P50 studies in patients with schizophrenia and relatives: differences in methodology between research groups. *Schizophr Res*. 2007;97:137–151.
43. de Wilde OM, Bour LJ, Dingemans PM, Koelman JHTM, Linszen DH. Failure to find P50 suppression deficits in young first-episode patients with schizophrenia and clinically unaffected siblings. *Schizophr Bull*. 2007;33:1319–1323.
44. Arnfred SM, Chen AC, Glenthoj BY, Hemmingsen RP. Normal p50 gating in unmedicated schizophrenia outpatients. *Am J Psychiatry*. 2003;160:2236–2238.
45. Light GA, Geyer MA, Clementz BA, Cadenhead KS, Braff DL. Normal P50 suppression in schizophrenia patients treated with atypical antipsychotic medications. *Am J Psychiatry*. 2000;157:767–771.
46. Rosburg T, Boutros NN, Ford JM. Reduced auditory evoked potential component N100 in schizophrenia–a critical review. *Psychiatry Res*. 2008;161:259–274.
47. Thoma RJ, Hanlon FM, Moses SN, et al. Lateralization of auditory sensory gating and neuropsychological dysfunction in schizophrenia. *Am J Psychiatry*. 2003;160:1595–1605.