Psychological Medicine

cambridge.org/psm

Original Article

*Co-first authors.

Cite this article: Lizano P, Dhaliwal K, Lutz O, Mothi SS, Miewald J, Montrose D, Keshavan M (2020). Trajectory of neurological examination abnormalities in antipsychotic-naïve first-episode psychosis population: a 1 year follow-up study. *Psychological Medicine* **50**, 2057–2065. https://doi.org/10.1017/S0033291719002162

Received: 27 March 2019
Revised: 10 July 2019
Accepted: 1 August 2019
First published online: 27 August 2019

Key words:
First-episode psychosis; global functioning; response; soft neurological signs; schizophrenia

Author for correspondence:
Paulo Lizano, E-mail: lizanopl@gmail.com

CAMBRIDGE UNIVERSITY PRESS

Trajectory of neurological examination abnormalities in antipsychotic-naïve first-episode psychosis population: a 1 year follow-up study

Paulo Lizano[1,2,*], Kiranpreet Dhaliwal[1,*], Olivia Lutz[1], Suraj Sarvode Mothi[3], Jean Miewald[4], Debra Montrose[4] and Matcheri Keshavan[1,2]

[1]Department of Psychiatry, Beth Israel Deaconess Medical Center, Boston, MA, USA; [2]Department of Psychiatry, Harvard Medical School, Boston, MA, USA; [3]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA and [4]Western Psychiatric Institute and Clinic, University of Pittsburgh, Pittsburgh, PA, USA

Abstract

Background. Neurological Examination Abnormalities (NES) are quantified by measuring subtle, partially localizable (cerebello-thalamo-prefrontal cortical circuit) and heritable neurological signs comprising sensory integration, motor coordination and complex motor sequencing that are associated with first-episode psychosis (FEP). A few studies have evaluated NES longitudinally and as a predictor for diagnostic and response classification, but these studies have been confounded, underpowered and divergent. We examined (1) baseline and longitudinal NES differences between diagnostic and year 1 response groups; (2) if NES predicts diagnostic and response groups and (3) relationships between clinical variables and NES measures in antipsychotic-naïve FEP.

Methods. NES and clinical measures were obtained for FEP-schizophrenia (FEP-SZ, $n = 232$), FEP non-schizophrenia (FEP-NSZ, $n = 117$) and healthy controls (HC, $n = 204$). Response groups with >25% improvement in average year 1 positive and negative symptomatology scores were classified as responsive ($n = 97$) and <25% improvement as non-responsive ($n = 95$). Analysis of covariance, NES trajectory analysis and logistic regression models assessed diagnostic and response group differences. Baseline and longitudinal NES relationships with clinical variables were performed with Spearman correlations. Data were adjusted for age, sex, race, socioeconomic status and handedness.

Results. Cognitive perceptual (COGPER) score was better than repetitive motor (REPMOT) at differentiating FEP-SZ from FEP-NSZ and distinguishing responders from non-responders. We identified significant group-specific associations between COGPER and worse GAF, positive and negative symptomatology and some of these findings persisted at 1-year assessment.

Conclusion. NES are an easy to administer, bedside-elicited, endophenotypic measure and could be a cost-effective clinical tool in antipsychotic-naïve FEP.

Introduction

In the USA, ~83 per 100 000 adolescents and young adults will experience the first-episode psychosis (FEP) each year (Simon *et al.*, 2017) and generally experience positive, negative and neurocognitive symptoms (Kahn *et al.*, 2015). The outcome of FEP can range from complete recovery to the development of schizophrenia (SZ) or non-schizophrenia (NSZ) resulting in serious functional impairments that are determined by neurocognitive and negative symptom severity (Reichenberg *et al.*, 2014). Advances in structural and functional neuroimaging have increased our understanding of the pathogenesis of SZ with functionally abnormal information processing in FEP and chronic SZ (Kahn *et al.*, 2015). However, a cost-effective clinical tool that can easily serve as a proxy for brain dysfunction, as well as aid in predicting diagnostic and treatment specificity in FEP is crucial in psychiatry.

Neurological examination abnormalities (NES) are quantified by measuring subtle, partially localizable (cerebello-thalamo-prefrontal cortical circuit) and heritable neurological signs comprising sensory integration, motor coordination and motor sequencing of complex movements (Keshavan *et al.*, 2003*b*; Bachmann *et al.*, 2014; Zhao *et al.*, 2014; Li *et al.*, 2018). Neuroscientific evidence for the importance of NES comes from the identification of cerebral correlates, covariation with cognition and its potential as a candidate endophenotypes for schizophrenia-spectrum disorders (Chan and Gottesman, 2008; Chan *et al.*, 2010). Moreover, a lifespan study by Chan *et al.* (2016) profiling NES scores across the schizophrenia spectrum and healthy controls has demonstrated that while healthy controls exhibit a U-shaped pattern for NES scores and age, patients on the schizophrenia spectrum

demonstrated a relatively stable elevation in NES scores (Chan *et al.*, 2016). Impairments in NES (higher scores) have been described in high-risk subjects (relatives of SZ patients, unaffected monozygotic twins discordant for SZ) (Torrey *et al.*, 1994; Prasad *et al.*, 2009; Bachmann *et al.*, 2014), antipsychotic-naïve FEP (Sanders *et al.*, 1994), chronic SZ patients (Sanders and Keshavan, 1998; Bachmann and Schröder, 2018) and other psychiatric diagnoses (Bombin, 2005; Chan *et al.*, 2016). However, the diagnostic specificity of NES has not been fully elucidated. In the largest cross-sectional study of NES in SZ spectrum disorders, it was shown that NES captures a moderate portion of psychosis proneness with reasonable specificity (Chan *et al.*, 2016). In a 6-month follow-up study of FEP, we have previously shown that factor scores for cognitively demanding and perceptual tasks were higher in the FEP schizophrenia (FEP-SZ) group relative to FEP non-schizophrenia (FEP-NSZ, participants with psychosis and a non-schizophrenia diagnosis) and control groups, but not between the FEP-NSZ and control group (Keshavan *et al.*, 2003*b*). Factor scores for baseline repetitive motor task abnormalities were elevated in both patient groups compared to controls, but not between the FEP groups, while factors scores for cognitive/perceptual tasks distinguished FEP-SZ from FEP-NSZ and HC groups, but not between FEP-NSZ and HC (Keshavan *et al.*, 2003*b*). Overall, few studies have evaluated the clinical utility of NES measures in predicting whether FEP patients develop either FEP-SZ or FEP-NSZ.

NES assessment could have the additional benefit of predicting psychosis symptom improvement. Treatment of FEP targets various domains, including positive and negative symptoms, cognitive dysfunction, social, academic and vocational functioning. In the literature, 'response' is defined as the reduction of total symptoms compared with baseline by 20–25% (minimally improved), 40–50% (much improved) whereas remission requires sustained mild positive, negative and disorganized symptoms for at least 6 months (Andreasen *et al.*, 2005). A meta-analysis of NES by Bachmann *et al.* (2014) and Bachmann and Schröder (2018) demonstrated that NES reductions (predominantly motor system subscales and to a lesser degree sensory integration scales) are more pronounced in patients with a remitting than in those with a non-remitting schizophrenia course, and that reductions of psychopathological symptoms paralleled the reduction in NES scores over time (Bachmann *et al.*, 2014; Bachmann and Schröder, 2018). Seven studies in this meta-analysis reported on remission status; however, the included studies were underpowered, did not include an antipsychotic-naïve FEP population, and the definition for remission varied with only one study using the Andreasen criteria (Prikryl *et al.*, 2012), while none used percent response categorization. Additionally, since antipsychotic effects on NES are measured indirectly, there have been mixed results regarding the side effects of antipsychotics on NES (Bachmann and Schröder, 2018), as well as mixed associations between NES and specific antipsychotic treatment response (Dazzan and Murray, 2002). To our knowledge, only a few small studies have evaluated the ability of baseline and follow-up scores of NES to predict response.

Taken together, an easy to administer, bedside-elicited, endophenotypic measure such as NES could act as a predictive clinical marker for differentiating FEP patients and response outcome. Thus, in the largest longitudinal study to date of NES we aim to study the role of NES in differentiating diagnostic and response groups in a prospective FEP population with minimal to no confounding influence of previous antipsychotic usage, chronicity and institutionalization. We hypothesize that NES will be able to (1) differentiate between FEP diagnostic groups and response outcome, (2) predict diagnostic and response group classification and (3) will be associated with psychopathology and functioning, but not antipsychotic or illness duration.

Methods

Participants

The study protocol and consent form were reviewed and approved by the IRB at the University of Pittsburgh and all subjects provided written informed consent or assent. The recruitment and assessment methods were described previously (Keshavan *et al.*, 2003*a*). From 1996 to 2004, the study population comprised of FEP patients from inpatient and outpatient services of the Western Psychiatric Institute and Clinic, Pittsburgh and they were under the care of Dr Keshavan. Patients were eligible to enter the study if they were aged 12–50 years, met the DSM-IV criteria for a psychotic disorder and had a 1-year follow-up assessment. Patient exclusion criteria included subjects with significant head injury, substance abuse or dependence, neurological/medical illness, prior antipsychotic exposure or mental retardation (Keshavan *et al.*, 2003*a*). All diagnoses were formally confirmed after at least 12 months of follow-up. Healthy comparison (HC) subjects were recruited from local neighborhoods and communities in which the patients resided, underwent a Structured Clinical Interview for DSM-IV–Non-Patient Edition (First *et al.*, 1997), and exclusion criteria included a current or previous axis I disorder, history of neurological or chronic medical problem with the potential to influence brain function, prior exposure to any psychotropic medication within 6 months of baseline assessment, first-degree relative history of schizophrenia or mood disorders or mental retardation (IQ < 75) (Keshavan *et al.*, 2003*a*). We did not include any subjects who were missing predominant hand information.

Our final baseline sample included 553 subjects (349 FEP probands and 204 HC subjects). FEP probands were divided based on their DSM-IV diagnosis into two groups: 232 FEP-SZ (participants with a schizophrenia $n = 65$, schizophreniform $n = 10$, schizoaffective $n = 44$ or residual/unspecified SZ diagnosis $n = 113$) and 117 FEP-NSZ (participants with psychosis and non-schizophrenia diagnosis, such as Bipolar 1 disorder $n = 17$, major depressive affective disorder $n = 33$ or delusional disorder $n = 12$, reactive psychosis not otherwise specified or psychosis not otherwise specified $n = 55$). For a breakdown of diagnosis by maximum time point, see online Supplementary Table 1. For a comparison of baseline demographic and clinical differences between FEP participants that followed-up ($n = 194$) *v.* those that dropped-out ($n = 155$) at 1-year, see online Supplementary Table 2.

Clinical assessment

Neurological evaluations were carried out using an inter-rater reliable modified version of the Buchanan–Heinrich Neurological Evaluation Scale (Sanders *et al.*, 1998) at several time points (baseline, week 4, week 8, week 26 and year 1). All of the NES measurements were conducted by a trained and reliable rater who was blind to the clinical data and to the subjects' groups, had consistently adequate reliability (a detailed description and evaluation of inter-rater reliabilities and the purpose for modifying the NES battery can be found in the following study) (Sanders

https://doi.org/10.1017/S0033291719002162 Published online by Cambridge University Press

et al., 1998) and baseline evaluations were performed prior to treatment with antipsychotics. For the diagnostic group analysis, week 4 data were excluded since only two HC subjects had NES data at this time point. Prior factor analysis and principal component analysis data (Keshavan *et al.*, 2003*b*) yielded two factors: repetitive motor (REPMOT; fist-ring, fist-edge-palm, alternating fist-palm and rapid alternating movements) and cognitive perceptual (COGPER, audiovisual integration, face-hand test and verbal memory). The average total NES score (TOT13) was also reported (Buchanan and Heinrichs, 1989; Keshavan *et al.*, 2003*b*). To test the stability of these two factors, we repeated the factor analysis performed by Keshavan *et al.* (2003*a*, *b*) using the same 13 NES scores on our larger sample and we obtained the same two clusters (data not shown). For a breakdown (mean and standard deviation) of REPMOT and COGPER sub-scores by diagnosis and by time, see online Supplementary Table 3. From the 553 subjects assessed at baseline, 56 subjects were missing baseline REPMOT, COGPER, and TOT13 scores [32 FEP-SZ (13.8%), 14 FEP-NSZ (12%) and 10 HC (4.9%)]. The Amelia II R package was used for modal imputations ($n = 100$) using race, sex and group as nominal variables, and this was performed in all three groups at baseline (Honaker *et al.*, 2011). For subjects with NES data at baseline and another follow-up time point, we calculated a NES change score by simple subtraction.

Duration of untreated psychosis (DUP) was defined as the number of weeks between the onset of psychotic symptoms and index admission into this study. This was determined by consensus based on SCID, medical records and review by the diagnostic group chaired by senior clinicians (MK or DM). Average IQ was collected from all patients using Ammon's quick IQ test (Ammons and Ammons, 1962). Global functioning measures were obtained from the global assessment of functioning scale (GAF) (Endicott *et al.*, 1976) and average positive and negative symptoms were obtained from the Scale for the Assessment of Positive Symptoms (SAPS) and Negative Symptoms (SANS) (Andreasen, 1990), respectively. Assessment of extrapyramidal symptoms was performed using items addressing bradykinesia-rigidity, tremor and akathisia (McEvoy *et al.*, 1991).

Missing baseline Hollingshead Four-Factor Index socioeconomic status (Hollingshead, 1975) data were imputed for both missing mother ($n = 48$, 8.7%) and father ($n = 68$, 12%) data as described above, followed by averaging the parents socioeconomic status score.

Response classification

A subset of FEP probands with baseline and 1-year follow-up data for total symptoms (e.g. SANS and SAPS scores) were characterized into response and non-response groups at 1-year based on definitions reviewed by Kahn *et al.* (2015). We were unable to use the Schizophrenia Working Group definition for remission in schizophrenia because of the relatively short follow-up of our naturalistic FEP study design, as well as the required strict remission criteria of 6 months of sustained symptom remission, which led to an underpowered sample size for subsequent analyses ($n = 1$ remitter at year 1) (Andreasen *et al.*, 2005). This observation was likely due to our naturalistic study design, which meant that some participants followed up at 6-months and some didn't, which made it difficult to obtain the remission status for many of our subjects. Response groups were created by calculating a percent change score (year 1-baseline/baseline) and participants with greater than a 25% reduction in average symptoms (combined SAPS and SANS) from baseline were characterized as responsive ($n = 99$) and those with less than 25% reduction as non-responsive ($n = 95$). A 25% improvement threshold was chosen to enhance the sample size of the group comparison.

Statistical analyses

All statistical analyses were performed using the R statistical analysis software (version 3.3.3, https://www.r-project.org/). Baseline sociodemographic and clinical differences between groups were analyzed with independent chi-squared tests or one-way analysis of variance (ANOVA). NES and clinical measures were assessed for normality utilizing visual inspection of histograms. We have >95% power to detect a 1.13 mean difference between FEP-SZ and FEP-NSZ, with a standard deviation of 1.3 and sigma of 0.05 (Keshavan *et al.*, 2003*b*).

A repeated measures ANOVA was used to assess both a group (diagnostic and treatment response), time, group by time interaction effect and group by response by time interaction effect on REPMOT, COGPER and TOT13 using random (subject) and fixed effects (age of consent, sex, race, parental socioeconomic status, handedness and antipsychotic status). Post-hoc pair-wise contrasts (FEP-NSZ to HC, FEP-SZ to HC, FEP-NSZ to FEP- SZ, Response to Non-Response) were run for REPMOT, COGPER, TOT13, SAPS, SANS and GAF utilizing a general linear model at each time point using data adjusted for age at consent, sex, race, parental socioeconomic status, handedness and antipsychotic status. To test the effects of IQ on NES we performed an additional analysis including IQ as a covariate.

Logistic regression analysis was used to examine a baseline univariate and multivariate prediction model with diagnostic group classification as the dependent variable and predictive variables being baseline REPMOT, COGPER and TOT13 measures. Crude and adjusted odds ratios (OR), Akaike information criteria (AIC) and 95% confidence intervals were calculated for each model. A similar approach was taken for response prediction with response status as the dependent variable and predictive variables being baseline and change from baseline for week 4, week 8 and week 26 REPMOT, COGPER and TOT13 measures. All diagnostic group data was co-varied for age at consent, sex, race and handedness, but not antipsychotic status since this did have a significant contribution to our model in the repeated measure and post-hoc analysis. Response data were additionally co-varied for year 1 antipsychotic usage to consider the effect of treatment. The area under the receiver operating curve (AUC) was used to assess the capacity of the index to distinguish diagnostic and response group comparisons.

Baseline and year 1 correlation between each NES measure and clinical measures for probands, FEP-SZ, FEP-NSZ, responders and non-responders were performed using Spearman correlations. False Discovery Rate (FDR) was used to correct p values for multiple comparisons.

Results

Demographics

Baseline sociodemographic and clinical information for diagnostic and response groups are summarized in Table 1. Sex and socioeconomic status were significantly different across diagnostic groups, while race and handedness were significantly different in the response groups. At baseline, the FEP-SZ group showed

https://doi.org/10.1017/S0033291719002162 Published online by Cambridge University Press

Table 1. Baseline demographic information for diagnostic and response groups

		Diagnostic group			Response group at year 1		
	HC (*N* = 204)	FEP-SZ (*N* = 232)	FEP-NSZ (*N* = 117)	*p*-Value	Responsive (*N* = 99)	Non-responsive (*N* = 95)	*p*-Value
Sex (female/male)	102/102	78/154	49/68	0.002	31/68	33/62	0.723
Race (AA/OT/CA)	61/16/127	92/16/124	34/9/74	0.190	20/10/69	41/5/49	0.002
Handedness (L/M/R)	13/5/186	21/9/202	5/9/103	0.094	8/9/82	8/1/86	0.040
Age, mean (s.d.)	24.3 (6.9)	25.5 (8.3)	23.7 (7.7)	0.077	23.3 (8.0)	25.5 (7.8)	0.054
SES, mean (s.d.)	42.7 (11.5)	38.5 (12.8)	42.7 (12.1)	<0.001	42.0 (13.7)	38.5 (12.0)	0.058
Average IQ, mean (s.d.)	107.4 (7.5)	99.5 (13.4)	102.8 (14.5)	<0.001	102.5 (12.7)	101.5 (14.0)	0.614
SAPS, mean (s.d.)	–	0.9 (0.5)	0.5 (0.4)	<0.001	1.0 (0.6)	0.6 (0.4)	<0.001
SANS, mean (s.d.)	–	2.2 (0.6)	1.8 (0.4)	<0.001	2.2 (0.6)	1.9 (0.5)	<0.001
Average SAPS and SANS, mean (s.d.)	–	1.5 (0.4)	1.2 (0.3)	<0.001	1.6 (0.5)	1.3 (0.3)	<0.001
GAF, mean (s.d.)	–	32.8 (9.8)	38.2 (10.7)	<0.001	32.5 (10.9)	36.6 (10.5)	0.009
DUP, mean (s.d.)	–	167.6 (285.6)	95.1 (199.6)	0.015	86.8 (181.2)	130.9 (196.7)	0.108
Year 1 antipsychotic status (Yes/No)	–	94/50	27/37	0.004	63/36	56/39	0.601

FEP, first-episode psychosis; SZ, schizophrenia; NSZ, non-schizophrenia; HC, healthy controls; AA, African American; OT, other; CA, Caucasian; L, left; M, mixed; R, right; SES, socioeconomic status; SAPS, Scale for the Assessment of Positive Symptoms; SANS, Scale for the Assessment of Negative Symptoms; GAF, global assessment of functioning; DUP, duration of untreated psychosis.

Table 2. Repeated measures ANOVA for group, time, group by time and response by diagnosis by time interaction

		Group		Time		Group by Time		Response by Diagnosis by Time	
		F-value	*p*-value	*F*-value	*p*-value	*F*-value	*p*-value	*F*-value	*p*-value
Diagnostic groups	REPMOT	52.41	<0.001	4.12	0.003	1.13	0.337	–	–
	COGPER	119.59	<0.001	12.61	<0.001	2.45	0.012	–	–
	TOT13	132.01	<0.001	8.25	<0.001	2.05	0.037	–	–
Response groups	REPMOT	12.79	<0.001	4.44	0.002	0.23	0.924	0.82	0.513
	COGPER	0.03	0.855	8.84	<0.001	2.74	0.028	1.00	0.406
	TOT13	8.58	0.004	7.09	<0.001	0.56	0.691	0.22	0.926

ANOVA, analysis of variance; REPMOT, repetitive motor; COGPER, cognitive perceptual, TOT13: average neurological evaluation scale total score. Co-varied for age, sex, race, handedness, socioeconomic status and antipsychotic status. Results were similar when excluding antipsychotic status as a covariate.

significantly worse IQ, positive and negative symptoms, GAF scores, longer duration of illness and greater year 1 antipsychotic use compared to the FEP-NSZ group (208 FEP-SZ and NSZ participants had year 1 data). With regards to response classification, the response group demonstrated worse positive and negative symptoms, and GAF scores at baseline. There were no significant differences for IQ, duration of illness or year 1 antipsychotic status between response groups. There were no significant differences for race, handedness, age, socioeconomic class, IQ, psychosis severity or GAF between the drop-out and follow-up groups, but there was a greater percentage of women and a longer DUP in the drop-out group (online Supplementary Table 2).

Clinical course of psychosis

In the diagnostic group analysis, there was a significant group, time and group by time interaction ($p < 0.01$) for all NES measures, except for group by time in REPMOT (Table 2). Race had a significant moderating effect on REPMOT ($p = 0.013$, $F = 3.18$), but not for COGPER or TOT13. COGPER, REPMOT and TOT13 measures decreased over the clinical course of psychosis (Fig. 1*a*) which corresponds to a reduction in psychopathological symptoms (online Supplementary Fig. 1*a*) and functional improvement (online Supplementary Fig. 2*a*). Both FEP groups demonstrated significantly greater NES measures across the factored groups and at all-time points compared to controls, with FEP-SZ demonstrating the greatest impairments (Fig. 1*a*). Compared to FEP-NSZ, the FEP-SZ group had significantly worse COGPER and TOT13 scores at baseline and week 26 (Fig. 1*a*). Additionally, the FEP-SZ group demonstrated worse SANS, SAPS, average psychopathology (online Supplementary Fig. 1*a*) and GAF scores (online Supplementary Fig. 2*a*) compared to FEP-NSZ subjects at all-time points analyzed, except for year 1 SAPS, which demonstrated a trending difference.

https://doi.org/10.1017/S0033291719002162 Published online by Cambridge University Press

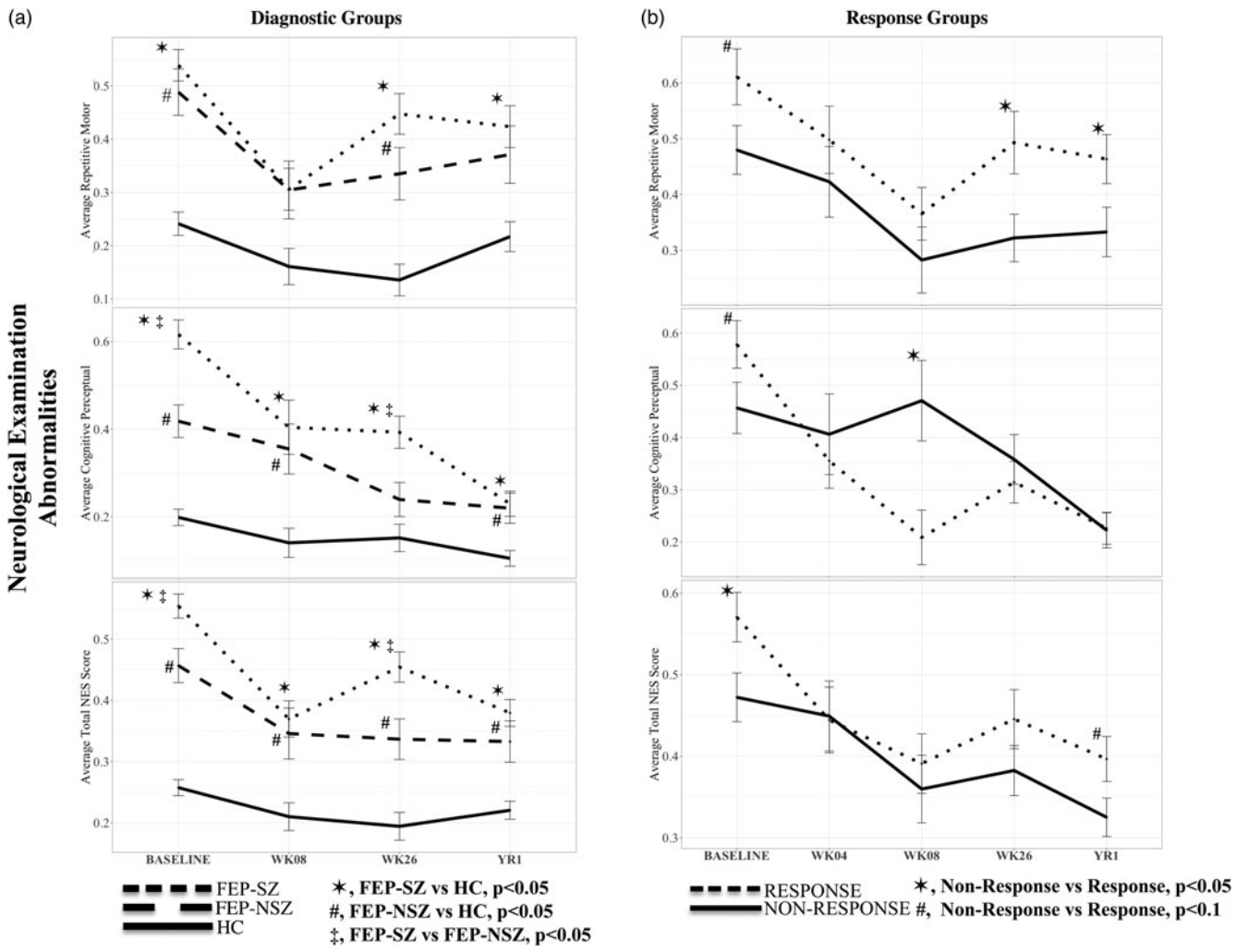

Fig. 1. Longitudinal changes in repetitive motor, cognitive perceptual, average total NES score with standard error bars across (a) diagnostic groups (b) response classification at year 1. FEP-SZ, first-episode psychosis schizophrenia; FEP-NSZ, first-episode psychosis non-schizophrenia; HC, healthy control; NES: neurological evaluation scale. Adjusted for age, sex, race, handedness and socioeconomic status.

Response groups displayed a significant group effect for REPMOT and TOT13 ($p < 0.005$), as well as a significant time effect for all three measures ($p < 0.003$), and a significant group by time effect for COGPER ($p = 0.025$) (Table 2). There was no significant group by response by time effect on any of the NES measures (Table 2). Race did not play a significant moderating effect on response status for any of the NES measures. The response group showed that NES measures decreased over time (Fig. 1*b*), which corresponds to symptomatic (online Supplemental Fig. 1*b*) and functional (online Supplementary Fig. 2*b*) improvement. Compared to non-responders, the response group demonstrated a significantly greater difference for baseline TOT13 ($p < 0.05$) and a trending increase for REPMOT and COGPER ($p < 0.1$) (Fig. 1*b*). REPMOT was significantly increased in responders relative to non-responders at week 26 and year 1. At week 8, COGPER was significantly reduced in responders compared to non-responders (Fig. 1*b*). Responders also demonstrated significantly greater improvements in SAPS, SANS, average psychopathology (online Supplementary Fig. 1*b*) and GAF scores (online Supplementary Fig. 2*b*) at week 26 and year 1 compared to non-responders. Additionally, there was no interaction between antipsychotic status and NES scores between responders and non-responders (data not shown). Similar significant results were obtained for the diagnostic and response categorization when performing repeated measures ANOVA using unimputed data (data not shown). The distribution of psychosis severity change score in the response groups can be found in online Supplementary Fig. 3.

Prediction analysis

For the diagnostic group analysis, univariate and multivariate regressions demonstrated that baseline REPMOT, COGPER and TOT13 could predict FEP-SZ [AUC (SE): REPMOT = 0.710 (0.025), COGPER = 0.762 (0.023), TOT13 = 0.810 (0.021)] or FEP-NSZ [AUC (SE): REPMOT = 0.664 (0.033), COGPER = 0.663 (0.032), TOT13 = 0.711 (0.030)] classification compared to controls, with FEP-SZ demonstrating the greatest risk prediction (Table 3, online Supplementary Fig. 4*a*, *b*). Baseline COGPER (OR = 2.58, $p = 0.001$) and TOT13 (OR = 3.96, $p = 0.001$) were able to significantly differentiate between FEP-SZ and FEP-NSZ subjects [AUC (SE): REPMOT = 0.540 (0.033), COGPER = 0.624 (0.031), TOT13 = 0.611 (0.033)] even after controlling for socio-demographic variables (Table 3, online Supplementary Fig. 4*c*).

https://doi.org/10.1017/S0033291719002162 Published online by Cambridge University Press

Table 3. Logistic regression evaluating the effect of baseline neurological examination abnormalities on diagnostic group classification

	Mean (S.D.)			HC v. FEP-NSZ			HC v. FEP-SZ			FEP-SZ v. FEP-NSZ		
	FEP-SZ	FEP-NSZ	HC	OR crude (95% CI)	OR co-varied (95% CI)	*p*-Value	OR crude (95% CI)	OR co-varied (95% CI)	*p*-Value	OR crude (95% CI)	OR co-varied (95% CI)	*p*-Value
Repetitive motor	0.54 (0.45)	0.49 (0.48)	0.24 (0.31)	5.39 (2.8–10.4)	5.90 (2.9–11.7)	<0.001	10.43 (5.6–19.5)	9.11 (4.8–17.3)	<0.001	1.45 (0.9–2.4)	1.36 (0.8–2.3)	0.251
Cognitive perceptual	0.62 (0.50)	0.42 (0.40)	0.20 (0.27)	6.53 (3.2–13.4)	8.50 (3.9–18.4)	<0.001	22.57 (11.5–44.4)	22.03 (10.8–44.9)	<0.001	3.02 (1.8–5.1)	2.58 (1.5–4.5)	0.001
Average total NES score	0.55 (0.30)	0.46 (0.30)	0.26 (0.19)	45.82 (14–156)	54.07 (15–192)	<0.001	368.24 (111–1217)	337.32 (99–1156)	<0.001	4.27 (1.9–9.6)	3.96 (1.7–9.2)	0.001

HC, healthy controls; FEP-NSZ, first-episode psychosis non-schizophrenia; FEP-SZ, first-episode psychosis schizophrenia; OR, odds ratio; CI, confidence interval; NES, neurological evaluation scale. Co-varied for age, sex, race, handedness, socioeconomic status.

In the response group analysis, univariate and multivariate regressions demonstrated that baseline REPMOT (OR = 2.01, $p = 0.048$), COGPER (OR = 2.06, $p = 0.033$) and TOT13 (OR = 4.01, $p = 0.012$) could significantly predict response classification [AUC (SE): REPMOT = 0.576 (0.041), COGPER = 0.598 (0.041), TOT13 = 0.591 (0.041)] compared to non-responders (Table 4, online Supplementary Fig. 5). When accounting for DUP, the results remained significant for COGPER ($p = 0.042$) and TOT13 ($p = 0.017$), but not for REPMOT ($p = 0.055$). Baseline psychosis severity scores were not included in the model, since baseline scores are accounted for in the response classification, making these two variables collinear. Furthermore, changes in COGPER at week 4 (OR = 0.34, $p = 0.037$) and week 8 (OR = 0.10, $p = 0.016$) from baseline predicted response compared to non-response (Table 4). When co-varying for year 1 antipsychotic usage, baseline NES measures still predicted response classification. Additionally, there was no significant interaction effect between baseline NES scores and year 1 antipsychotic usage (data not shown). Univariate effect of diagnostic group classification, DUP or year 1 antipsychotic usage status did not predict responders from non-responders (data not shown), and as a result DUP was not included in the multivariate analysis.

NES covariance with clinical measures

In the baseline correlational analysis, TOT13 demonstrated the greatest relationship with other clinical measures and this effect was mostly driven by COGPER (Fig. 2*a*). Specifically, higher COGPER scores were significantly associated with (1) worse GAF scores in probands and FEP-SZ, (2) greater SAPS scores in probands, (3) worse SANS in probands, FEP-SZ and non-responders and (4) greater average psychopathology scores in probands and FEP-SZ groups (Fig. 2*a*). No baseline relationships were observed for REPMOT except for GAF in probands (Fig. 2*a*). In a post hoc analysis, we examined the relationship between extrapyramidal symptoms (EPS) and NES scores at baseline and year 1. At baseline, none of the participants was on antipsychotics, but there was a significant positive correlation between EPS and REPMOT ($r = 0.210$, $p = 0.005$), COGPER ($r = 0.196$, $p = 0.009$) and TOT13 ($r = 0.253$, $p < 0.001$). At year 1, there was a significant positive relationship between EPS score with REPMOT ($r = 0.419$, $p < 0.001$) and TOT13 ($r = 0.393$, $p < 0.001$), but not COGPER ($r = 0.122$, $p = 0.254$) while controlling for antipsychotic status. Lastly, while there was a significant negative relationship between IQ and TOT13 ($r = -0.443$, $p \leqslant 0.001$), REPMOT ($r = -0.284$, $p \leqslant 0.001$) and COGPER ($r = -0.402$, $p \leqslant 0.001$) at baseline, the group comparisons remained similar after additionally controlling for IQ (data not shown).

In the year 1 correlational analysis, COGPER continued to demonstrate the greatest number of relationships with clinical measures (Fig. 2*b*). Specifically, COGPER was significantly ($p < 0.05$, corrected) correlated with greater SANS and average psychopathology scores in responders. No REPMOT relationships were observed at year 1.

Discussion

In summary, we found that both the diagnostic and response groups demonstrated a pattern in which NES measures for COGPER, REPMOT and TOT13 decreased over the clinical course of psychosis paralleling a reduction in psychopathological symptoms and functional improvement, with FEP-SZ and

https://doi.org/10.1017/S0033291719002162 Published online by Cambridge University Press

Table 4. Logistic regression evaluating the effect of neurological examination abnormalities on response at year 1.

	Response		Non-response				
	N	Mean (s.d.)	*N*	Mean (s.d.)	OR-crude (95% CI)	OR-co-varied (95% CI)	*p*-Value
Repetitive motor							
Baseline	99	0.61 (0.50)	95	0.48 (0.42)	1.63 (0.88–3.03)	2.01 (1.01–4.0)	0.048
Week 4 Δ	57	−0.14 (0.46)	40	−0.06 (0.41)	0.67 (0.26–1.75)	0.56 (0.19–1.66)	0.295
Week 8 Δ	26	−0.16 (0.39)	32	−0.13 (0.34)	0.75 (0.18–3.18)	1.29 (0.22–7.70)	0.780
Week 26 Δ	73	−0.11 (0.52)	71	−0.17 (0.43)	1.31 (0.65 -2.62)	1.18 (0.53–2.62)	0.684
Cognitive perceptual							
Baseline	99	0.58 (0.45)	95	0.46 (0.05)	1.29 (0.72–2.29)	2.06 (1.06–4.0)	0.033
Week 4 Δ	57	−0.23 (0.47)	40	−0.04 (0.47)	0.41 (0.16–1.03)	0.34 (0.12–0.94)	0.037
Week 8 Δ	26	−0.34 (0.44)	32	0.09 (0.45)	0.11 (0.03–0.45)	0.10 (0.02–0.66)	0.016
Week 26 Δ	73	−0.21 (0.40)	71	−0.16 (0.52)	0.81 (0.40–1.65)	0.57 (0.25–1.29)	0.175
Average total NES score							
Baseline	99	0.57 (0.30)	95	0.47 (0.29)	1.98 (0.78-5.06)	4.01 (1.36–11.86)	0.012
Week 4 Δ	57	−0.14 (0.28)	40	−0.04 (0.21)	0.20 (0.4–1.08)	0.17 (0.03–1.12)	0.065
Week 8 Δ	26	−0.13 (0.22)	32	−0.4 (0.22)	0.15 (0.01–1.75)	0.34 (0.01–8.20)	0.505
Week 26 Δ	73	−0.10 (0.26)	71	−0.12 (0.26)	1.46 (0.41–5.27)	1.05 (0.25–4.39)	0.942

NES, neurological evaluation scale; OR, odds ratio; CI, confidence interval, co-varied for age, sex, race, handedness, socioeconomic status and year 1 antipsychotic usage. Δ, change indicates change from given time point to baseline.

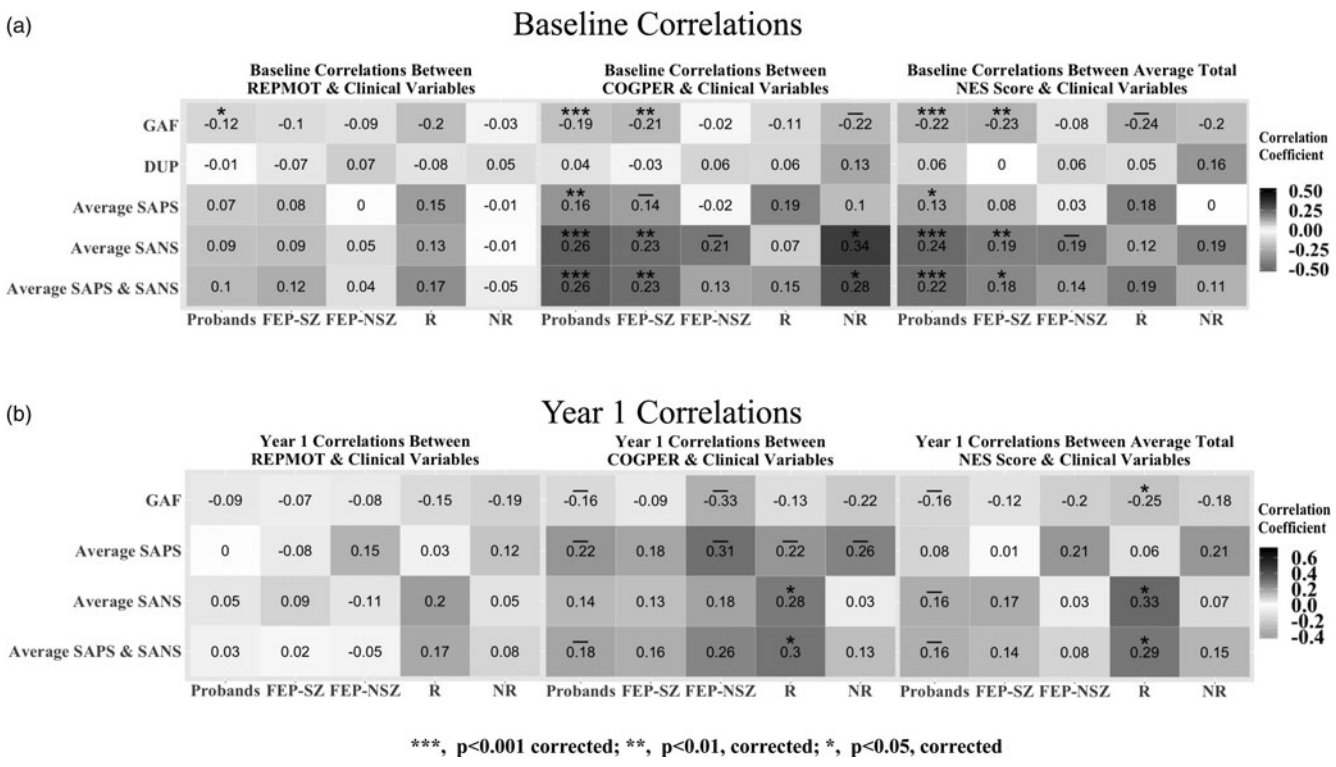

Fig. 2. Correlations between REPMOT, COGPER and total NES scores and clinical variables across diagnostic and response classification groups at (a) baseline and (b) year 1. REPMOT, repetitive motor; COGPER, cognitive perceptual; NES, Neurological Evaluation Scale, R, response; NR, non-response; GAF, Global Assessment Scale; DUP, duration of psychosis; SAPS, Scale for the Assessment of Positive Symptoms; SANS, Scale for the Assessment of Negative Symptoms. Data were adjusted for age, sex, race, handedness and socioeconomic status. (–), $p < 0.1$.

https://doi.org/10.1017/S0033291719002162 Published online by Cambridge University Press

non-responders exhibiting worse trajectories. Additionally, we found that baseline TOT13, and its subscale COGPER, were better than REPMOT at (1) predicting psychosis compared to controls, (2) differentiating FEP-SZ from FEP-NSZ and (3) potentially distinguishing responders from non-responders (this effect was not driven by either FEP-SZ or FEP-NSZ groups, nor by DUP). Additionally, changes in COGPER at week 4 and week 8 from baseline differentiated responders from non-responders, whereas REPMOT and TOT13 did not. We identified significant group-specific associations between COGPER and worse GAF, positive and negative symptomatology and some of these findings persisted at year 1 assessment. Similar to our repeated measures ANOVA and logistic regression analysis, we found that better COGPER scores at baseline and year 1 were associated with better psychosis symptoms (specifically SAPS and average SAPS/SANS). Lastly, we showed that neither diagnostic group classification, DUP or year 1 antipsychotic usage status had an effect on response classification. These observations are of clinical relevance since NES's may be a prognostic/predictive biomarker in an antipsychotic-naïve FEP population.

The findings reported here are consistent with the literature, which suggests that while NES scores decrease in the clinical course of schizophrenia with improvement of psychopathological symptoms, they remain impaired compared to controls (Bachmann *et al*., 2014). We expanded on this body of literature by examining a large group of FEP-NSZ and found that while they have fewer impairments compared to the FEP-SZ group, their clinical trajectories are similar, an observation which has not been previously described. Additionally, group differences reported here replicate our past work, which showed that compared to HC and FEP-NSZ, FEP-SZ patients demonstrated higher NES scores with COGPER scores being markedly worse (Keshavan *et al*., 2003*b*). We showed that there is a significant positive relationship between EPS and NES scores at baseline and year 1 (even after controlling for year 1 antipsychotic use), which could be due to an overlap in symptoms, but further work is needed to elucidate this relationship. We also showed that there is a significant negative relationship between IQ and NES scores at baseline which did not affect our group comparison results and is consistent with the literature (Chan *et al*., 2016). In a large cross-sectional study of NES across the psychosis spectrum showed that in their other psychiatric disorders group (21% bipolar disorder and 18% major depression disorder) there were no significant NES differences but did not comment on the group's psychosis status (Chan *et al*., 2016). While our findings parallel our earlier work, this study is unique in that we included a larger sample and performed a predictive analysis, which demonstrated that baseline COGPER assessment may be utilized to differentiate FEP groups, which can have important implications for monitoring disease progression or to identify subjects with an increased liability toward schizophrenia. We also showed longitudinal NES stability in HCs, which is consistent with a developmental longitudinal study of 5-year duration, demonstrating that in late childhood, children typically have elevated NES scores that gradually decline with motor maturation into young adulthood and are not related to learning effects (Martins *et al*., 2008). In our HC sample of young adults, we observed low NES scores that remained stable over time, and the latter could not be explained by learning effects.

Consistent with the response literature comparing NES (Bachmann *et al*., 2014), we found that baseline NES and psychosis severity was significantly higher in responders (determined by >25% improvement in positive and negative symptom scores) and that a greater decrease in NES across the clinical course was more evident in responders *v.* non-responders. The non-responders tended to have less NES and psychosis severity compared to responders at baseline, but the rate of change was less prominent in the non-responder group and this was independent of DUP. A few studies have reported increased NES in non-remitters compared to remitters at the follow-up time point, but these studies have been limited by small sample sizes, inclusion of FEP participants already on antipsychotics, and inconsistent definition of remission status (Bachmann *et al*., 2014). A 4-year follow-up study of FEP used the Andreasen remission criteria of schizophrenia and demonstrated that remitters had improved total NES score and sensory integration/motor sequencing, while non-remitters had worse total NES score over time, however, baseline NES could not differentiate these groups (Prikryl *et al*., 2012). The association between the degree of change in NES and response categorization is consistent with previous findings (Bachmann *et al*., 2014; Bachmann and Schröder, 2018). We extended this area of research by demonstrating that baseline COGPER and its repeated assessment at week 4 and week 8 might be viable course predictors in antipsychotic-naïve FEP. Additionally, we accounted for the effect of diagnostic group classification, DUP or year 1 antipsychotic usage and did not identify these as predictors of response in our sample, despite divergent findings in the literature (Carbon and Correll, 2014). Our study is distinctive in that we performed the largest naturalistic study to date of antipsychotic-naïve FEP patients demonstrating the effectiveness as a bedside clinical tool.

NES are generally present prior to SZ diagnosis and treatment initiation suggesting that NES are an intrinsic feature of SZ (Keshavan *et al*., 2003*b*; Bachmann *et al*., 2005). Also, findings from structural and functional MRI studies support the conceptualization of NES as a manifestation of the cerebello-thalamo-prefrontal brain network model of schizophrenia disorders and related psychotic disorders with some of these regions (activation in cortical motor areas, the thalamus and the cerebellum) potentially being heritable in mono- and di-zygotic twins (Zhao *et al*., 2014; Li *et al*., 2018). Thus, NES trajectories coupled with other diagnostic criteria may facilitate earlier SZ diagnoses and thereby improve prognosis. Since an extended duration of untreated illness is generally associated with a worse prognosis (Loebel *et al*., 1992), earlier recognition of these signs can lead to earlier diagnosis and earlier exposure to preventative measures. While our findings did not show significant relationships between NES and DUP, we found that COGPER was a better predictor of diagnostic and response classification (independent of diagnostic group). Additionally, the positive associations between COGPER and symptomology, predominantly with negative symptoms, but positive symptoms as well, are consistent with previous literature (Bombin, 2005; Chan *et al*., 2015). Moreover, past literature has not examined the relationship between NES and clinical variables across the clinical course and here we showed that there is stability between COGPER impairment and worse clinical outcomes. Therefore, NES changes seem to parallel symptomology and evaluating NES trajectories may help predict response.

There are a number of strengths of this FEP study including evaluating NES without the confounding influences of chronicity or previous antipsychotic usage. By determining the effects of NES over time, longitudinal studies show variable patterns better than when compared to cross-sectional studies. Furthermore, to our knowledge, this is the largest sample size of a single site

https://doi.org/10.1017/S0033291719002162 Published online by Cambridge University Press

longitudinal study reporting NES scores. We acknowledge some limitations to this study, including patient attrition, relatively short duration of follow-up, a limited collection of cognitive domains and lack of comprehensive antipsychotic dosage information. Notwithstanding these limitations, our study points to the value of further research in NES as a potentially valuable bedside marker of treatment response.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291719002162

Acknowledgements. The authors thank the families who took part in this study, and the many participants who contributed to this project. This publication was supported by funds received from National Institute of Mental Health grants (NIMH) MH-45203, MH-01180 and MH- 45156 (MSK) and by NIH General Clinical Research Center grant M01 RR-00056. We thank Kevin Eklund RN MSN who performed the neurological assessments.

References

Ammons RB and Ammons CH (1962) The quick test (QT): provisional manual. *Psychological Reports* **11**, 111–161.

Andreasen NC (1990) Methods for assessing positive and negative symptoms. *Modern Problems of Pharmacopsychiatry* **24**, 73–88.

Andreasen NC, Carpenter WT, Kane JM, Lasser RA, Marder SR and Weinberger DR (2005) Remission in schizophrenia: proposed criteria and rationale for consensus. *American Journal of Psychiatry* **162**, 441–449.

Bachmann S and Schröder J (2018) Neurological soft signs in schizophrenia: an update on the state- versus trait-perspective. *Frontiers in Psychiatry* **8**, 453.

Bachmann S, Bottmer C and Schröder J (2005) Neurological soft signs in first-episode schizophrenia: a follow-up study. *American Journal of Psychiatry* **162**, 2337–2343.

Bachmann S, Degen C, Geider FJ and Schröder J (2014) Neurological soft signs in the clinical course of schizophrenia: results of a meta-analysis. *Frontiers in Psychiatry* **5**, 703.

Bombin I (2005) Significance and meaning of neurological signs in schizophrenia: two decades later. *Schizophrenia Bulletin* **31**, 962–977.

Buchanan RW and Heinrichs DW (1989) The neurological evaluation scale (NES): a structured instrument for the assessment of neurological signs in schizophrenia. *Psychiatry Research* **27**, 335–350.

Carbon M and Correll CU (2014) Clinical predictors of therapeutic response to antipsychotics in schizophrenia. *Dialogues in Clinical Neuroscience* **16**, 505–524.

Chan RCK and Gottesman II (2008) Neurological soft signs as candidate endophenotypes for schizophrenia: a shooting star or a Northern star? *Neuroscience and Biobehavioral Reviews* **32**, 957–971.

Chan RCK, Xu T, Heinrichs RW, Yu Y and Wang Y (2010) Neurological soft signs in schizophrenia: a meta-analysis. *Schizophrenia Bulletin* **36**, 1089–1104.

Chan RCK, Geng F-L, Lui SSY, Wang Y, Ho KKY, Hung KSY, Gur RE, Gur RC and Cheung EFC (2015) Course of neurological soft signs in first-episode schizophrenia: relationship with negative symptoms and cognitive performances. *Scientific Reports* **5**, 11053.

Chan RCK, Xie W, Geng F-L, Wang Y, Lui SSY, Wang C-Y, Yu X, Cheung EFC and Rosenthal R (2016) Clinical utility and lifespan profiling of neurological soft signs in schizophrenia spectrum disorders. *Schizophrenia Bulletin* **42**, 560–570.

Dazzan P and Murray RM (2002) Neurological soft signs in first-episode psychosis: a systematic review. *The British Journal of Psychiatry. Supplement* **43**, s50–s57.

Endicott J, Spitzer RL, Fleiss JL and Cohen J (1976) The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry* **33**, 766–771.

First MB, Spitzer RL, Williams JB, Gibbon M, First (1997). User's guide for the structured clinical inverview for DSM-IV axis I disorders SCID-1: clinician version. American Psychiatric Pub.

Hollingshead AB (1975). Four factor index of social status. Unpublished manuscript, Yale University, NewHaven, CT.

Honaker J, King G and Blackwell M (2011) Amelia II: a program for missing data. *Journal of Statistical Software* **45**, 1–47.

Kahn RS, Sommer IE, Murray RM, Meyer-Lindenberg A, Weinberger DR, Cannon TD, O'Donovan M, Correll CU, Kane JM, Van Os J and Insel TR (2015) Schizophrenia. *Nature Reviews. Disease Primers* **1**, 15067.

Keshavan MS, Haas G, Miewald J, Montrose DM, Reddy R, Schooler NR and Sweeney JA (2003*a*) Prolonged untreated illness duration from prodromal onset predicts outcome in first episode psychoses. *Schizophrenia Bulletin* **29**, 757–769.

Keshavan MS, Sanders RD, Sweeney JA, Diwadkar VA, Goldstein G, Pettegrew JW and Schooler NR (2003*b*) Diagnostic specificity and neuroanatomical validity of neurological abnormalities in first-episode psychoses. *American Journal of Psychiatry* **160**, 1298–1304.

Li Z, Huang J, Xu T, Wang Y, Li K, Zeng Y-W, Lui SSY, Cheung EFC, Jin Z, Dazzan P, Glahn DC and Chan RCK (2018) Neural mechanism and heritability of complex motor sequence and audiovisual integration: a healthy twin study. *Human Brain Mapping* **39**, 1438–1448.

Loebel AD, Lieberman JA, Alvir JM, Mayerhoff DI, Geisler SH and Szymanski SR (1992) Duration of psychosis and outcome in first-episode schizophrenia. *American Journal of Psychiatry* **149**, 1183–1188.

Martins I, Lauterbach M, Slade P, Luís H, DeRouen T, Martin M, Caldas A, Leitão J, Rosenbaum G and Townes B (2008) A longitudinal study of neurological soft signs from late childhood into early adulthood. *Developmental Medicine & Child Neurology* **50**, 602–607.

McEvoy JP, Hogarty GE and Steingard S (1991) Optimal dose of neuroleptic in acute schizophrenia: a controlled study of the neuroleptic threshold and higher haloperidol dose. *Archives of General Psychiatry* **48**, 739–745.

Prasad KM, Sanders R, Sweeney J, Montrose D, Diwadkar V, Dworakowski D, Miewald J and Keshavan M (2009) Neurological abnormalities among offspring of persons with schizophrenia: relation to premorbid psychopathology. *Schizophrenia Research* **108**, 163–169.

Prikryl R, Ceskova E, Tronerova S, Kasparek T, Kucerova HP, Ustohal L, Venclikova S and Vrzalova M (2012) Dynamics of neurological soft signs and its relationship to clinical course in patients with first-episode schizophrenia. *Psychiatry Research* **200**, 67–72.

Reichenberg A, Feo C, Prestia D, Bowie CR, Patterson TL and Harvey PD (2014) The course and correlates of everyday functioning in schizophrenia. *Schizophrenia Research. Cognition* **1**, e47–e52.

Sanders RD and Keshavan MS (1998) The neurologic examination in adult psychiatry: from soft signs to hard science. *The Journal of Neuropsychiatry and Clinical Neurosciences* **10**, 395–404.

Sanders RD, Keshavan MS and Schooler NR (1994) Neurological examination abnormalities in neuroleptic-naive patients with first-break schizophrenia: preliminary results. *American Journal of Psychiatry* **151**, 1231–1233.

Sanders RD, Forman SD, Pierri JN, Baker RW, Kelley ME, Van Kammen DP and Keshavan MS (1998) Inter-rater reliability of the neurological examination in schizophrenia. *Schizophrenia Research* **29**, 287–292.

Simon GE, Coleman KJ, Yarborough BJ, Operskalski B, Stewart C, Hunkeler E, Lynch FL, Carrell D and Beck A (2017) Incidence and presentation of first-episode psychosis in a population-based sample. *Psychiatric Services* **68**, 456–461.

Torrey EF, Taylor EH, Bracha HS, Bowler AE, McNeil TF, Rawlings RR, Quinn PO, Bigelow LB, Rickler K and Sjostrom K (1994) Prenatal origin of schizophrenia in a subgroup of discordant monozygotic twins. *Schizophrenia Bulletin* **20**, 423–432.

Zhao Q, Li Z, Huang J, Yan C, Dazzan P, Pantelis C, Cheung EFC, Lui SSY and Chan RCK (2014) Neurological soft signs are not 'soft' in brain structure and functional networks: evidence from ALE meta-analysis. *Schizophrenia Bulletin* **40**, 626–641.