International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 23(1): 25–35 (2014)
Published online 21 January 2014 in Wiley Online Library
(wileyonlinelibrary.com) **DOI:** 10.1002/mpr.1429

Multilevel ordinal factor analysis of the Positive and Negative Syndrome Scale (PANSS)

JAN STOCHL,[1,2] PETER B. JONES,[1,2] JAMES PLAISTOW,[2] ULRICH REININGHAUS,[3] STEFAN PRIEBE,[4] JESUS PEREZ[1,2] & TIM J. CROUDACE[1]

1 Department of Psychiatry, University of Cambridge, Cambridge, UK
2 Cameo Early Intervention Services, Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, UK
3 Section of Social Psychiatry, Health Service and Population Research Department, King's College London, Institute of Psychiatry, London, UK
4 Queen Mary University of London, Unit for Social and Community Psychiatry, Barts and the London School of Medicine, London, UK

Key words
psychometrics, psychoses, scale validation

Correspondence
Jan Stochl, Department of Psychiatry, University of Cambridge, Addenbrookes Hospital, Box 189, Cambridge, CB2 0QQ, UK.
Telephone: (+44) 7587146299
Email: js883@cam.ac.uk

Received 6 June 2012;
revised 6 September 2012;
accepted 6 November 2012

Abstract

Clinical assessments of the presence and severity of psychopathology are often collected by health care professionals in mental health services or clinical researchers trained to use semi-structured interviews. Clustering by interviewer or rater needs to be considered when performing psychometric analyses such as factor analysis or item response modelling as non-independence of observations arises in these situations. We apply more suitable multilevel methods to analyse ordinally scored Positive and Negative Syndrome Scale (PANSS) items. Our aim is to highlight the differences in results that occur when the data are analysed using a hierarchically sensitive approach rather than using a traditional (aggregated) analysis. Our sample ($n = 507$) consisted of patients diagnosed with schizophrenia who participated in a multi-centre randomized control clinical trial, the DIALOG study. Analyses reported and compared include an exploratory factor analysis as well as several recently published multifactor models re-estimated within a confirmatory analysis framework. Our results show that the fit of the model and the parsimony of the exploratory factor analysis (EFA) models indicated by the number of factors necessary to explain the inter-correlation among PANSS items improved significantly when data clustering is taken into account through multilevel analysis. Our modeling results support the pentagonal PANSS model first proposed by White *et al.* (1997). *Copyright © 2014 John Wiley & Sons, Ltd.*

Introduction

Definition and measurement of information concerning the mental state of individuals are of fundamental importance in psychiatry. They are central concerns for philosophers, clinicians and psychometricians, largely because much of the information involves subjective experience and is, essentially, unverifiable. Nevertheless, many measurement tools have been developed. These measurement tools are often administered by interviewers or raters rather than being completed by the patient (Ecob *et al.*, 2004). The Positive and Negative Symptom Scale (PANSS) is an example of a scale in which interviewers rate the severity of patient symptoms. Typically, a single rater assesses multiple patients, whether the rater is a clinician who provides treatment or a researcher who rates patients specifically for research purposes. Although interviewers are thoroughly trained to give the most objective rating possible, their ratings may still depend on their experience, subjective perception of the patient's symptom severity or other subjective judgments. Even if high interrater reliability is reported for a scale of this type, part of the variability between patients can still be attributed to variability between the raters.

The situation described earlier leads to some degree of correlation among the individual symptoms of patients who are assessed by the same rater, such that patients evaluated by the same rater may appear to be more similar to each other than they are to other patients. Traditional factor analysis methods are grounded on the assumption that the data being analysed come from independent and identically distributed observations (Bentler and Chou, 1987; Hox, 1993). This crucial assumption is obviously violated in the case of psychiatric data of this type, and the results (factor loadings, standard errors, *p*-values and item threshold parameters) of confirmatory factor analysis (CFA) or exploratory factor analysis (EFA) may therefore be biased (Julian, 2001; Muthén and Satorra, 1995).

Recent psychometric developments have attempted to address such variability in clustered data and hierarchical (also known as multilevel) modelling approaches have been developed for this purpose. Although originally developed for a variety of regression analyses, these techniques quickly became popular in latent variable applications, including traditional factor analysis of continuously scored items, but also categorical responses, i.e. item response theory methods. The popularity is boosted by the growing amount of software with capabilities to analyse latent variable models for hierarchical data, including a mixture of binary, ordinal, nominal and continuous item responses. The most popular packages include Mplus (Muthén and Muthén, 1998–2012), LISREL 8.80 (Jöreskog and Sörbom, 2007), or GLLAMM (Rabe-Hesketh *et al.*, 2004) in Stata.

If psychiatric datasets contain information on which clinician or rater assessed each patient, these methods provide a potentially effective and statistically well-motivated approach to preventing impact of inter-correlation (clustering) effects and thereby more accurately describe the latent structure of the data compared to traditional methods based on aggregated data.

Unfortunately, hierarchical modelling techniques have not yet been adopted in psychiatric research. To the best of our knowledge only 36 publications have used hierarchical regression and only two studies have employed hierarchical covariance structure modelling techniques in the top 10 rated psychiatric journals.

The aim of this study is to provide evidence for biased factor analytic estimates when clustering in the data is ignored. We present the results for an important psychopathology rating scale, the PANSS, which is widely used to assess the severity of psychotic symptoms. Our interest is in evaluating competing CFA models, using a more theoretically well motivated multilevel ordinal factor analysis model.

Factor analysis of Positive and Negative Syndrome Scale (PANSS)

The PANSS was originated as a rigorously operationalized method for evaluating positive, negative, and other symptom dimensions of schizophrenia (Kay, 1991). Since its publication in 1987 (Kay *et al.*, 1987) it has become one of the most widely used scales in mental health and has been successfully applied to other types of psychotic disorders beyond schizophrenia.

Many studies of the latent structure of psychopathology in psychosis investigate the construct validity of the PANSS using principal components analysis or factor analysis. Most existing studies have reported a five-factor structure, usually containing positive, negative, anxiety/depression/preoccupation, cognitive/disorganization/dysphoric and activation/excitement factors (Bell *et al.*, 1994a; Dollfus and Petit, 1995; Emsley *et al.*, 2003; Fitzgerald *et al.*, 2003; Lancon *et al.*, 1998; Lancon *et al.*, 2000; Lindenmayer *et al.*, 1994b; Lindenmayer *et al.*, 1995; Lykouras *et al.*, 2000; Wolthaus *et al.*, 2000). However, the items that load on each factor vary considerably from study to study. Other reported structures include two factors (positive, negative) (Kay *et al.*, 1987), three factors (positive, negative, relational/disorganization) (Peralta *et al.*, 1992; Strauss *et al.*, 1974), four factors (positive, negative, excitement/relational, depression/disorganization) (Kay and Sevy, 1990; Peralta *et al.*, 1994; Peralta *et al.*, 1992; Strauss *et al.*, 1974), six factors (the five factors described earlier, plus withdrawal) (Van den Oord *et al.*, 2006), seven factors (positive, negative, disorganization, excitement, depression, anxiety, motor) (Emsley *et al.*, 2003) and even eight factors (positive, negative, disorganization, excitement, depression, anxiety, preoccupation, somatization) (Peralta and Cuesta, 1994). Overall, the structure of the PANSS appears to be sample specific, and it is difficult to find an adequately fitting model or to replicate the structures found in existing studies (Van der Gaag *et al.*, 2006; White *et al.*, 1997). This unsatisfactory state of affairs is captured in summary form in Table 1. The values in Table 1 represent the number of studies in which the item was found to load on the corresponding factor.

There are two possible reasons for this extraordinary variability in the results of factor analytic studies. First, there may be a real lack of construct validity for this multidimensional psychopathology scale. Given its popularity and the other studies supporting its validity (Bell *et al.*, 1994b; El Yazaji *et al.*, 2002; Hatton *et al.*, 2005; Preston and Harrison, 2003; Van den Oord *et al.*, 2006), however, we consider this possibility to be unlikely. Second, none of the published studies have considered that PANSS is administered by clinical raters/interviewers, and often, at least a number of different clinicians would have collected the data. Although raters were well trained, there is still room for individual rating differences (for example, some PANSS raters may systematically overestimate or underestimate symptom severity).

Methods

Sample description

The sample ($n = 507$) of patients diagnosed with schizophrenia or a related psychotic disorder was obtained from a multi-centre randomized control trial, the DIALOG

Table 1. The PANSS factor loadings seen in previous factor analytic and principal component analytic studies

		positive	negative	anxietydepression	cognitive	excitement	disorganized	anxiety	preoccupied	others	delusionalor hallucinatory	withdrawal	thoughtdisorder	somatization	relational
Delusions	P1	**28**		1							1				
Conceptual disorganization	P2	7	1		**11**	1			1		1				
Hallucinatory behaviour	P3	**25**		1					1		1				
Excitement	P4	1		1		**26**	3	1					1		
Grandiosity	P5	**21**				2	1	1						1	
Suspiciousness/persecution	P6	**18**		3		3									
Hostility	P7	1		1		**24**					1				
Blunted affect	N1	1	**28**	1											
Emotional withdrawal	N2		**26**	2								1			3
Poor rapport	N3		**29**			2									
Social withdrawal	N4		**25**	2								1			3
Difficulty in abstract thinking	N5		4		**11**		7		1	1			1		
Lack of spontaneity	N6		**29**												
Stereotyped thinking	N7	3	5		4	3	**7**		1						
Somatic concern	G1	3		**8**				2						1	
Anxiety	G2	1		**19**				3							
Guilt feelings	G3			**21**			1								
Tension	G4			7	1	**13**	1	3							
Mannerisms & posturing	G5		1		**5**	2	**5**		1						
Depression	G6			**22**			1								
Motor retardation	G7		**2**	1											
Uncooperativeness	G8		2			**19**									
Unusual thought content	G9	**19**				1	2		1	1					
Disorientation	G10	1	3		**7**		6						1		
Poor attention	G11		2		9		**12**		1				1		
Lack of judgement & insight	G12	**13**	1		2	4	2				1				
Disturbance of volition	G13		**9**		2	1	3		1						
Poor impulse control	G14					**21**	1								
Preoccupation	G15	3	**5**	4	2	1	3		2						
Active social avoidance	G16		**15**	4		1	1				1	1			

study (Priebe *et al.*, 2007), that was designed to test a new computer-mediated intervention. The study was conducted in community mental health service sites in London (UK), Granada (Spain), Groningen (The Netherlands), Lund (Sweden), Mannheim (Germany), and Zurich (Switzerland) using the following inclusion criteria for patients: (1) living in the community and treated as outpatients (≥ three months of continuous care); (2) ICD-10 criteria for schizophrenia or related psychotic disorder (F20–F29); (3) aged between 18 and 65 years; (4) no organic or substance misuse disorder; (5) at least one meeting every two months with keyworker; (6) expectation to continue with the service for the next 12 months (Priebe *et al.*, 2007). Psychiatric diagnosis was obtained through a standardized and computer-based method using operationalized criteria (OPCRIT; McGuffin, 1991). The sample consists of 336 (66.3%) males and 171 (33.7%) females, with a mean age of 42.2 years [standard deviation (SD) = 11.4 years] and a mean illness duration of 15.9 years (SD = 10.3 years). The distribution of diagnoses was as follows: schizophrenia, 354 (69.8%); schizoaffective disorder, 73 (14.4%); delusional disorder, 3 (0.59%); and psychotic disorder not otherwise specified (NOS), 77 (15.2%). Patients were moderately ill when judged in terms of summary scores on standard rating scales for psychotic symptoms (PANSS mean score = 63.6, SD = 18.7).

Analysis

First, intraclass correlation (ICC) coefficients were computed to determine the magnitude of clustering. Then, traditional ordinal and multilevel ordinal EFAs were used to analyse the dataset, and the results of these methodologies were compared. Specifically, we compared number of factors and fit indices. Next, we employed and compared traditional ordinal CFA and multilevel ordinal CFA. We analysed several recently published PANSS factor structure models. We focused only on models that include the majority of the items; as a result, some models that focus only on a reduced number of PANSS items had to be excluded from consideration (Lindenmayer *et al.*, 1994a; Peralta *et al.*, 1994; Peralta *et al.*, 1992; Strauss *et al.*, 1974). The fit indices of the tested models are presented and compared. The best fitting model is subsequently presented in more detail.

In the multilevel factoring approach, the covariance matrix for the PANSS items was separated into between-level and within-level covariance matrices. The between-level covariance matrix is of little interest in our study for the following reasons: (a) there are only 13 raters in the DIALOG study, making the effective sample sizes for a between-level covariance matrix $n = 13$; and (b) construct validity of PANSS items essentially exists at the within-level. Thus, the factor structure was modelled using the within-level covariance matrix only. All analyses were estimated using MPlus software version 6.12 (Muthén and Muthén, 1998–2012). The PANSS items were specified as categorical ordered (ordinal) and a weighted least squares mean- and variance-adjusted (WLSMV) estimator was used for both the multilevel and traditional factor analysis to achieve ordinally sensitive analyses.

Results

Intraclass correlations (ICCs)

Table 2 shows the ICC coefficients for the DIALOG study PANSS data. An ICC can be interpreted as the proportion of between-level variance compared to the total variance of an item. A commonly adopted rule of thumb suggests that ICCs below 0.05 may indicate relatively minor clustering that can be ignored, allowing the data to be pooled.

The ICCs for the PANSS items in Table 2 are mostly higher than 0.05, which suggests that there is considerable variance between the raters that could distort the results of a traditional (non-multilevel) analysis. Multilevel factor analysis would therefore appear to be a more appropriate methodology than traditional psychometric approaches.

Exploratory factor analysis (EFA)

Modern software packages provide a number of fit indices for EFA models. Table 3 shows the fit indices when clustering is ignored and when it is acknowledged. An overall examination of the fit indices shows that the model fitting the data consists of six factors when the clustering is ignored. However, when clustering is considered, the four-factor solution seems to fit the data well, which is important from the perspective of parsimonious clinical interpretation. The factor loadings for this four-factor PANSS solution are shown in Table 4.

A more detailed examination of the four-factor structure shows that factor 1 (“negative”) contains five of the original negative symptoms (blunted affect, emotional withdrawal, poor rapport, social withdrawal, and lack of spontaneity), motor retardation (G7) and active social avoidance (G16). Factor 2 (“arousal”) contains eight of the original general psychopathology items (anxiety, tension, mannerisms and posturing, disorientation, poor attention, disturbance of volition, poor impulse control, and preoccupation), two positive items (excitement and conceptual disorganization) and one original negative item (stereotyped thinking). Factor 3 (“positive”) contains

Table 2. The intraclass correlation (ICC) coefficients of PANSS items

Variable		ICC	Variable		ICC
Delusions	P1	0.04	Anxiety	G2	0.09
Conceptual disorganization	P2	0.12	Guilt feelings	G3	0.09
Hallucinatory behaviour	P3	0.01	Tension	G4	0.15
Excitement	P4	0.10	Mannerisms & posturing	G5	0.31
Grandiosity	P5	0.08	Depression	G6	0.13
Suspiciousness/persecution	P6	0.07	Motor retardation	G7	0.20
Hostility	P7	0.26	Uncooperativeness	G8	0.18
Blunted affect	N1	0.13	Unusual thought content	G9	0.11
Emotional withdrawal	N2	0.19	Disorientation	G10	0.24
Poor rapport	N3	0.13	Poor attention	G11	0.15
Social withdrawal	N4	0.11	Judgement & insight	G12	0.26
Difficulty in abstract thinking	N5	0.03	Disturbance of volition	G13	0.10
Lack of spontaneity	N6	0.04	Poor impulse control	G14	0.28
Stereotyped thinking	N7	0.17	Preoccupation	G15	0.24
Somatic concern	G1	0.16	Active social avoidance	G16	0.17

four of the original positive symptoms (delusions, hallucinatory behaviour, suspiciousness/persecution, and grandiosity) and one general psychopathology item (unusual thought content. Finally, the "depression" factor consists of five original general psychopathology items (anxiety, guilt, tension, depression, and lack of judgment/insight).

Confirmatory factor analysis (CFA)

This section discusses several CFA models of the PANSS data previously published for psychotic populations. After extracting and specifying more than 20 alternative models from the literature, we compare the fit indices of the CFA models when data clustering is ignored or considered. Published models that include a smaller number of PANSS items, such as Peralta *et al.* (1992); Strauss *et al.* (1974); Kay and Sevy (1990); Peralta *et al.* (1994), and Lindenmayer *et al.* (1994a), were not estimated.

The results in Table 5 reveal that none of the models fit the data acceptably when the clustering is ignored. We observe considerable improvement in the fit indices for all of the models when modelling considered hierarchical aspects of the data. Few models fit the data relatively well, including the models of Emsley *et al.* (2003); Levine and Rabinowitz (2007); Reininghaus *et al.* (2013); Lancon *et al.* (1998) and White *et al.* (1997). However, the cutoff criteria for an acceptable fit are met only for one, the pentagonal factor model first proposed by White *et al.* (1997). Based on these results, we are forced to conclude that this model fits our data best, although some of the others may be close or conceptually quite similar. The factor loadings for the supported model are displayed in Table 6.

Discussion

Psychiatric research and mental health services make great use of interviewer rated assessments to determine severity of multiple dimensions of psychopathology in patients with mental disorders. The data from these tools require special approaches when analysed using factor analysis or item response theory. This study aimed to highlight the differences in results that occur when the data are analysed using the correct multilevel, hierarchically sensitive approach rather than using a traditional (aggregated) analysis.

The DIALOG study included the identities of the PANSS raters for each patient allowing us to compare the traditional approach (i.e. ignoring clustering) and the correct multilevel approach (i.e. using the available clustering information). In terms of model fit recently available for EFA models in modern statistical packages, fewer factors were necessary to satisfactorily explain inter-item correlations in terms of fit indices when the clustering was taken into account. Findings from a previous simulation study suggest that eigenvalues tend to be overestimated when the clustering is ignored and other fit indices also show poorer fit of the model (Stochl, 2012).

Although the fit of the model to the data is important, we also acknowledge that substantial clinical interpretation of the factors beyond the statistical testing of the factor solution is essential. Four factors denoted as "negative", "positive",

Table 3. The exploratory factor analysis (EFA) fit indices when clustering is ignored and acknowledged

Number of factors	Chi-square		df		*p*-Value		RMSEA		CFI		TLI		SRMR	
	I	A	I	A	I	A	I	A	I	A	I	A	I	A
1	3544	1601	405	810	<0.001	<0.001	0.124	0.044	0.693	0.693	0.670	0.670	0.136	0.150
2	2300	1070	376	781	<0.001	<0.001	0.101	0.027	0.812	0.888	0.782	0.875	0.097	0.098
3	1627	858	348	753	<0.001	0.005	0.085	0.017	0.875	0.959	0.843	0.953	0.075	0.073
4	1109	749	321	726	<0.001	0.267	0.070	0.008	0.923	0.991	0.895	0.989	0.057	0.056
5	731	697	295	700	<0.001	0.523	0.054	0.000	0.957	1.000	0.937	1.001	0.042	0.046
6	550	656	270	675	<0.001	0.690	0.045	0.000	0.973	1.000	0.956	1.009	0.035	0.040
7	413	618	246	651	<0.001	0.820	0.037	0.000	0.984	1.000	0.971	1.017	0.028	0.033
8	315	587	223	628	<0.001	0.880	0.029	0.000	0.991	1.000	0.982	1.022	0.024	0.027

Note: RMSEA, root mean square error of approximation; CFI, comparative fit index; TLI, Tucker–Lewis index; SRMR, standardized root mean square residual; I, clustering ignored; A, clustering acknowledged.

"arousal" and "depression" that correspond quite closely to the clinical phenomenology in this cohort could be distinguished. The factor labelled "negative" contains five of the original negative symptoms (blunted affect, emotional withdrawal, poor rapport, social withdrawal, and lack of spontaneity) plus items motor retardation (G7) and active social avoidance (G16). This result is consistent with previous factor analytic studies, which have also loaded these items on the "negative" subscale (Bell *et al.*, 1994a; Fredrikson *et al.*, 1997; Lee *et al.*, 2003; Van den Oord *et al.*, 2006). This factor also has negative associations with both grandiosity and excitement, which may be consistent with an affective dimension.

Factor "arousal" contained eight of the original general psychopathology items (anxiety, tension, mannerisms and posturing, disorientation, poor attention, disturbance of volition, poor impulse control, and preoccupation), two positive items (excitement and conceptual disorganization) and one original negative item (stereotyped thinking). In previous five-factor solutions, these items have variously loaded upon the "excitement" and "cognitive" dimensions (Lancon *et al.*, 1998; Mass *et al.*, 2000). Anxiety (G2) and preoccupation (G15) have tended to load on the "depression" or "negative" dimensions in prior research (Van den Oord *et al.*, 2006). In this study, however, both are consistent with a factor indicating "arousal".

Factor "positive" contained four of the original positive symptoms (delusions, hallucinatory behaviour, suspiciousness/persecution, and grandiosity) and one general psychopathology item (unusual thought content); previous studies have consistently grouped these items with the positive symptoms (Emsley *et al.*, 2003; Lancon *et al.*, 1998; Wolthaus *et al.*, 2000).

The "depression" factor consisted of five original general psychopathology items (anxiety, guilt, tension, depression, and lack of judgment/insight). Previous factor analytic solutions have consistently loaded these items on a "depression" factor (Bell *et al.*, 1994b; Fredrikson *et al.*, 1997; Lee *et al.*, 2003; Van den Oord *et al.*, 2006). The association between increased insight and increased depression has been consistently replicated in this first-episode client group (Birchwood *et al.*, 2000).

A strength of this study was the systematic review of existing PANSS factor solutions detailed in Table 1. In the modelling phase, several recently published PANSS models were tested within an ordinally sensitive multilevel CFA framework. To our knowledge, all of the published models that have applied factor analysis to PANSS data have ignored the clustering of PANSS data by the raters (many have also treated item responses as continuous). For each candidate factor model, we showed the difference

Table 4. The factor loadings of the four-factor EFA model when clustering is considered

		Factor 1	Factor 2	Factor 3	Factor 4
Delusions	P1	0.04	−0.09	**0.96**	0.01
Conceptual disorganization	P2	0.01	**0.48**	0.22	−0.36
Hallucinatory behaviour	P3	−0.02	0.01	**0.73**	0.04
Excitement	P4	−**0.56**	**0.79**	−0.08	0.01
Grandiosity	P5	−**0.42**	0.12	**0.53**	−0.25
Suspiciousness/persecution	P6	0.14	0.15	**0.53**	0.14
Hostility	P7	−0.05	0.34	0.31	0.02
Blunted affect	N1	**0.78**	0.06	−0.04	−0.06
Emotional withdrawal	N2	**0.83**	0.01	0.01	0.12
Poor rapport	N3	**0.71**	0.09	−0.08	−0.16
Social withdrawal	N4	**0.73**	0.03	0.09	0.27
Difficulty in abstract thinking	N5	0.31	0.24	0.11	−0.26
Lack of spontaneity	N6	**0.79**	0.02	−0.15	−0.14
Stereotyped thinking	N7	0.01	**0.61**	0.14	−0.06
Somatic concern	G1	0.11	0.34	0.07	0.32
Anxiety	G2	0.00	**0.46**	0.10	**0.60**
Guilt feelings	G3	0.03	0.28	0.08	**0.41**
Tension	G4	−0.03	**0.71**	−0.15	**0.44**
Mannerisms & posturing	G5	0.21	**0.40**	0.03	−0.25
Depression	G6	0.20	0.31	−0.01	**0.51**
Motor retardation	G7	**0.77**	−0.05	−0.04	−0.11
Uncooperativeness	G8	0.20	0.37	0.01	−0.28
Unusual thought content	G9	−0.05	0.01	**0.84**	−0.19
Disorientation	G10	0.19	**0.43**	0.01	−0.17
Poor attention	G11	0.11	**0.52**	0.11	−0.27
Lack of judgement & insight	G12	0.09	0.07	0.33	−**0.58**
Disturbance of volition	G13	0.23	**0.43**	0.00	0.00
Poor impulse control	G14	−0.16	**0.61**	0.10	−0.02
Preoccupation	G15	0.30	**0.47**	0.15	0.09
Active social avoidance	G16	**0.55**	0.00	0.38	0.38

Note: Values higher than 0.4 in absolute magnitude are highlighted in bold typeface.

in fit when clustering is ignored versus acknowledged. For all models, we demonstrated that model fit increased when clustering was acknowledged.

Some limitations of these results must be outlined. We analysed the data using a two-level approach (patients nested within raters) although in reality additional levels of nesting are present in the data: (a) raters are nested within centres and (b) the study was conducted in two different languages (that is centres are nested within "languages"). Unfortunately, modelling higher than two-level data is not possible in the current implementation of the software. Further, model estimation used only one estimator (i.e. WLSMV). Therefore, findings require careful replication using other estimation methods (e.g. full-information maximum likelihood) before firm conclusions can be drawn. Finally, the factorial structure at the rater level was not considered in this study due to small number of raters; future studies with a considerable larger number of raters may help to model the relationships at the higher levels of nesting.

Despite the limitations of this study, however, we provided considerable systematic evidence of bias in the results for key parameter estimates when factor analysis of categorical data (i.e. item response modelling) is used for data analysis and the clustering of individuals within psychiatric data is ignored. In addition, we provided more accurate account of the factor structure of the widely used PANSS instrument.

The essential implication of this study is to encourage researchers or mental health services that aspire to collect successive outcome measures on patients, to identify clinical raters during data collection. When such data is subsequently processed, this identification allows

Int. J. Methods Psychiatr. Res. 23(1): 25–35 (2014). DOI: 10.1002/mpr

Table 5. The fit indices of the previously published models when clustering is ignored

Title	Factors	Chi-square		df		*p*-Value		CFI		TLI		RMSEA	
		I	A	I	A	I	A	I	A	I	A	I	A
Van den Oord *et al.* (2006)	4	4239	—	366	—	<0.001	—	0.609	—	0.567	—	0.145	—
Bell *et al.* (1994a)	5	5280	1817	376	782	<0.001	<0.001	0.480	0.611	0.439	0.596	0.161	0.051
Bell *et al.* (1994b)	5	5763	1747	376	782	<0.001	<0.001	0.464	0.641	0.422	0.627	0.169	0.05
Dollfus and Petit (1995)	5	6055	—	252	—	<0.001	—	0.328	—	0.264	—	0.214	—
Dudek (2005)	5	4647	1614	324	675	<0.001	<0.001	0.550	0.649	0.512	0.635	0.163	0.053
Emsley *et al.* (2003)	5	1920	1014	395	830	<0.001	<0.001	0.851	0.93	0.836	0.927	0.088	0.021
White *et al.* (1997)	5	1084	646	258	558	<0.001	0.006	0.896	0.961	0.880	0.958	0.080	0.018
Fredrikson *et al.* (1997)	5	4448	1577	299	624	<0.001	<0.001	0.542	0.619	0.502	0.603	0.166	0.055
Higashima *et al.* (1998)	5	3014	859	183	393	<0.001	<0.001	0.556	0.716	0.490	0.697	0.175	0.049
Lancon *et al.* (1998)	5	1456	687	242	518	<0.001	<0.001	0.850	0.928	0.829	0.923	0.100	0.025
Lancon *et al.* (1999)	5	5195	1652	400	835	<0.001	<0.001	0.530	0.682	0.489	0.668	0.154	0.044
Lancon *et al.* (2000)	5	3235	1186	251	527	<0.001	<0.001	0.643	0.726	0.607	0.713	0.154	0.05
Lee *et al.* (2003)	5	4773	—	374	—	<0.001	—	0.560	—	0.523	—	0.153	—
Levine and Rabinowitz (2007)	5	1984	1032	395	830	<0.001	<0.001	0.844	0.924	0.829	0.920	0.089	0.022
Lykouras *et al.* (2000)	5	3360	1257	252	528	<0.001	<0.001	0.608	0.702	0.570	0.688	0.157	0.052
Kay and Sevy (1990)	5	4761	1615	350	728	<0.001	<0.001	0.515	0.630	0.476	0.616	0.158	0.049
Marder *et al.* (1997)	5	5394	1820	403	838	<0.001	<0.001	0.511	0.616	0.472	0.602	0.157	0.048
Mass *et al.* (2000)	5	1346	611	199	430	<0.001	<0.001	0.841	0.917	0.815	0.911	0.107	0.029
Van den Oord *et al.* (2006)	5	3969	1413	365	771	<0.001	<0.001	0.630	0.725	0.588	0.711	0.140	0.041
Wolthaus *et al.* (2000)	5	1895	956	366	772	<0.001	<0.001	0.846	0.931	0.829	0.927	0.091	0.022
Fredrikson *et al.* (1997)	6	5519	1935	322	673	<0.001	<0.001	0.456	0.528	0.407	0.508	0.179	0.061
Van den Oord *et al.* (2006)	6	4265	—	367	—	<0.001	—	0.595	—	0.552	—	0.145	—
El Yazaji *et al.* (2002)	6	—	—	—	—	—	—	—	—	—	—	—	—
Reininghaus *et al.* (2013)	6	1726	965	375	810	<0.001	<0.001	0.868	0.941	0.846	0.937	0.085	0.020
Emsley *et al.* (2003)	7	2064	991	356	762	<0.001	<0.001	0.817	0.905	0.791	0.899	0.098	0.024
Peralta and Cuesta (1994)	8	2079	948	344	750	<0.001	<0.001	0.828	0.925	0.796	0.918	0.100	0.023

Note: RMSEA, root mean square error of approximation; CFI, comparative fit index; TLI, Tucker–Lewis Index; SRMR, standardized root mean square residual; I, clustering ignored; A, clustering acknowledged. The models with missing fit indices did not converge.

Table 6. The standardized factor loadings of the model proposed by White *et al.* (1997)

Negative	Item		Estimate	SE	*p*-value
Negative	Lack of spontaneity	N6	0.752	0.028	<0.001
	Blunted affect	N1	0.811	0.028	<0.001
	Emotional withdrawal	N2	0.809	0.027	<0.001
	Poor rapport	N3	0.771	0.036	<0.001
	Social withdrawal	N4	0.709	0.034	<0.001
	Motor retardation	G7	0.726	0.027	<0.001
	Mannerisms & posturing	G5	0.654	0.06	<0.001
	Uncooperativeness	G8	0.46	0.042	<0.001
	Disturbance of volition	G13	0.124	0.05	0.013
	Poor impulse control	G14	0.188	0.046	<0.001
Positive	Delusions	P1	0.888	0.013	<0.001
	Unusual thought content	G9	0.87	0.014	<0.001
	Grandiosity	P5	0.513	0.037	<0.001
	Hallucinatory behaviour	P3	0.707	0.035	<0.001
	Somatic concern	G1	0.111	0.032	0.001
Activation	Hostility	P7	0.769	0.07	<0.001
	Poor impulse control	G14	0.712	0.046	<0.001
	Excitement	P4	0.581	0.022	<0.001
	Uncooperativeness	G8	0.361	0.063	<0.001
	Poor rapport	N3	−0.142	0.048	0.003
	Tension	G4	0.283	0.046	<0.001
Dysphoric	Anxiety	G2	0.766	0.033	<0.001
	Tension	G4	0.582	0.05	<0.001
	Guilt feelings	G3	0.539	0.045	<0.001
	Depression	G6	0.559	0.029	<0.001
	Somatic concern	G1	0.489	0.051	<0.001
Autistic	Poor attention	G11	0.625	0.026	<0.001
	Preoccupation	G15	0.774	0.02	<0.001
	Difficulty in abstract thinking	N5	0.538	0.038	<0.001
	Stereotyped thinking	N7	0.679	0.027	<0.001
	Disturbance of volition	G13	0.473	0.053	<0.001
	Hallucinatory behaviour	P3	0.068	0.04	0.092

proper acknowledgement of clustering using appropriate analytical models.

Acknowledgements

PBJ acknowledges support from the Department of Health Policy Research Programme for the funding for the National Eden Study, and from the NIHR for the CLAHRC for Cambridgeshire and Peterborough. PBJ, JP, JS and TJC acknowledge funding support from the NIHR programme grant RP-PG-0606-1335. The DIALOG study was funded by the Research Directorate of the European Commission within the Framework Programme 5 (QLG5-CT-2002-01938).

References

Bell M.D., Lysaker P.H., Beam-Goulet J.L., Milstein R.M., Lindenmayer J.P. (1994a). Five-component model of schizophrenia: assessing the factorial invariance of the positive and negative syndrome scale. *Psychiatry Research*, **52**(3), 295–303.

Bell M.D., Lysaker P.H., Milstein R.M., Beam-Goulet J.L. (1994b) Concurrent validity of the cognitive component of schizophrenia: relationship of PANSS scores to neuropsychological assessments. *Psychiatry Research*, **54**(1), 51–58.

Bentler M.P., Chou C.P. (1987) Practical issues in structural modeling. *Sociological Methods and Research*, **16**(1), 78–117.

Birchwood M., Iqbal Z., Chadwick P., Trower P. (2000) Cognitive approach to depression and suicidal thinking in psychosis. *The British Journal of Psychiatry*, **177**(6), 516–528.

Dollfus S., Petit M. (1995) Principal-component analyses of PANSS and SANS-SAPS in schizophrenia: their stability in an acute phase. *European Psychiatry*, **10**(2), 97–106.

Dudek P.T. (2005). *An examination of the factor structure of the SCI-PANSS*. Philadelphia: Unpublished Doctoral thesis, Drexel University.

Ecob R., Croudace T.J., White I.R., Evans J.E., Harrison G.L., Sharp D. *et al.* (2004) Multilevel investigation of variation in HoNOS ratings by mental health professionals: a naturalistic study of consecutive referrals. *International Journal of Methods in Psychiatric Research*, **13**(3), 152–164.

El Yazaji M., Battas O., Agoub M., Moussaoui D., Gutknecht C., Dalery J. *et al.* (2002) Validity of the depressive dimension extracted from principal component analysis of the PANSS in drug-free patients with schizophrenia. *Schizophrenia Research*, **56**(1–2), 121–127.

Emsley R., Rabinowitz J., Torreman M. (2003) The factor structure for the Positive and Negative Syndrome Scale (PANSS) in recent-onset psychosis. *Schizophrenia Research*, **61**(1), 47–57.

Fitzgerald P.B., de Castella A.R., Brewer K., Filia K., Collins J., Davey P. *et al.* (2003) A confirmatory factor analytic evaluation of the pentagonal PANSS model. *Schizophrenia Research*, **61**(1), 97–104.

Fredrikson D.H., Steiger J.M., MacEwan G.W., Altman S., Kopala L.C., Flynn S.W. *et al.* (1997) PANSS symptom factors in schizophrenia. *Schizophrenia Research*, **24**(1–2), 15–15.

Hatton C., Haddock G., Taylor J.L., Coldwell J., Crossley R., Peckham N. (2005) The reliability and validity of general psychotic rating scales with people with mild and moderate intellectual disabilities: an empirical investigation. *Journal of Intellectual Disability Research*, **49**(Pt 7), 490–500.

Higashima M., Urata K., Kawasaki Y., Maeda Y., Sakai N., Mizukoshi C. *et al.* (1998) P300 and the thought disorder factor extracted by factor-analytic procedures in schizophrenia. *Biological Psychiatry*, **44**(2), 115–120.

Hox J.J. (1993) Factor analysis of multilevel data: gauging the Muthén method. In Oud J.H.L., van Blokland-Vogelesang R.A.W. (eds.), Advances in Longitudinal and Multivariate Analysis in the Behavioral Sciences, pp. 141–156. Nijmegen: ITS.

Jöreskog K.G., Sörbom D. (2007) LISREL (Version 8.80). Chicago, IL: Scientific Software International.

Julian M.W. (2001) The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, **8**(3), 325–352.

Kay S.R. (1991) Positive and Negative Syndromes in Schizophrenia: Assessment and Research. New York: Brunner/Mazel.

Kay S.R., Fiszbein A., Opler L.A. (1987) The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, **13**(2), 261–276.

Kay S.R., Sevy S. (1990) Pyramidical model of schizophrenia. *Schizophrenia Bulletin*, **16**(3), 537–545.

Lancon C., Aghababian V., Llorca P.M., Auquier P. (1998) Factorial structure of the Positive and Negative Syndrome Scale (PANSS): a forced five-dimensional factor analysis. *Acta Psychiatrica Scandinavica*, **98**(5), 369–376.

Lancon C., Auquier P., Nayt G., Reine G. (2000) Stability of the five-factor structure of the Positive and Negative Syndrome Scale (PANSS). *Schizophrenia Research*, **42**(3), 231–239.

Lancon C., Reine G., Llorca P.M., Auquier P. (1999) Validity and reliability of the French-language version of the Positive and Negative Syndrome Scale (PANSS). *Acta Psychiatrica Scandinavica*, **100**(3), 237–243.

Lee K.H., Harris A.W., Loughland C.M., Williams L.M. (2003) The five symptom dimensions and depression in schizophrenia. *Psychopathology*, **36**(5), 226–233.

Levine S.Z., Rabinowitz J. (2007) Revisiting the 5 dimensions of the Positive and Negative Syndrome Scale. *Journal of Clinical Psychopharmacology*, **27**(5), 431–436.

Lindenmayer J.P., Bernstein-Hyman R., Grochowski S. (1994a) Five-factor model of schizophrenia. Initial validation. *The Journal of Nervous and Mental Disease*, **182**(11), 631–638.

Lindenmayer J.P., Bernstein-Hyman R., Grochowski S. (1994b) A new five factor model of schizophrenia. *The Psychiatric Quarterly*, **65**(4), 299–322.

Lindenmayer J.P., Grochowski S., Hyman R.B. (1995) Five factor model of schizophrenia: replication across samples. *Schizophrenia Research*, **14**(3), 229–234.

Lykouras L., Oulis P., Psarros K., Daskalopoulou E., Botsis A., Christodoulou G.N. *et al.* (2000) Five-factor model of schizophrenic psychopathology: how valid is it? *European Archives of Psychiatry and Clinical Neuroscience*, **250**(2), 93–100.

Marder S.R., Davis J.M., Chouinard G. (1997) The effects of risperidone on the five dimensions of schizophrenia derived by factor analysis: combined results of the North American trials. *Journal of Clinical Psychiatry*, **58**(12), 538–546.

Mass R., Schoemig T., Hitschfeld K., Wall E., Haasen C. (2000) Psychopathological syndromes of schizophrenia: evaluation of the dimensional structure of the positive and negative syndrome scale. *Schizophrenia Bulletin*, **26**(1), 167–177.

McGuffin P.F.A.H.I. (1991) A polydiagnostic application of operational criteria in studies of psychotic illness: development and reliability of the opcrit system. *Archives of General Psychiatry*, **48**(8), 764–770.

Muthén B., Satorra A. (1995) Complex sample data in structural equation modeling. *Sociological Methodology*, **25**, 267–316.

Muthén L., Muthén B. (1998–2012) Mplus: Statistical Analysis with Latent Variables (Version 6.1). Los Angeles, CA: Muthén.

Peralta V., Cuesta M.J. (1994) Psychometric properties of the positive and negative syndrome scale (PANSS) in schizophrenia. *Psychiatry Research*, **53**(1), 31–40.

Peralta V., Cuesta M.J., de Leon J. (1994) An empirical analysis of latent structures underlying schizophrenic symptoms: a four-syndrome model. *Biological Psychiatry*, **36**(11), 726–736.

Peralta V., de Leon J., Cuesta M.J. (1992) Are there more than two syndromes in schizophrenia? A critique of the positive–negative dichotomy. *The British Journal of Psychiatry*, **161**, 335–343.

Preston N.J., Harrison T.J. (2003) The brief symptom inventory and the Positive and Negative Syndrome Scale: discriminate validity between a self-reported and observational measure of psychopathology. *Comprehensive Psychiatry*, **44**(3), 220–226.

Priebe S., McCabe R., Bullenkamp J., Hansson L., Lauber C., Martinez-Leal R. *et al.* (2007) Structured patient–clinician communication and 1-year outcome in community mental healthcare: cluster randomised controlled trial. *The British Journal of Psychiatry*, **191**, 420–426.

Rabe-Hesketh S., Skrondal A., Pickles A. (2004) Generalized multilevel structural equation modelling. *Psychometrika*, **69**(2), 167–190.

Reininghaus U., Priebe S., Bentall R.P. (2013) Testing the psychopathology of psychosis: evidence for a general psychosis dimension. *Schizophrenia Bulletin*, **39**(4), 884–895.

Stochl J. (2012) Effects of Ignoring Clustered Data Structures in Factor Analysis with Applications to Psychiatry. Cambridge: University of Cambridge.

Strauss J.S., Carpenter Jr W.T., Bartko J.J. (1974) The diagnosis and understanding of schizophrenia. Part III. Speculations on the processes that underlie schizophrenic symptoms and signs. *Schizophrenia Bulletin*, **11**, 61–69.

Van den Oord E.J., Rujescu D., Robles J.R., Giegling I., Birrell C., Bukszar J. *et al.* (2006) Factor structure and external validity of the PANSS revisited. *Schizophrenia Research*, **82**(2–3), 213–223.

Van der Gaag M., Cuijpers A., Hoffman T., Remijsen M., Hijman R., de Haan L. *et al.* (2006) The five-factor model of the Positive and Negative Syndrome Scale I: confirmatory factor analysis fails to confirm 25 published five-factor solutions. *Schizophrenia Research*, **85**(1–3), 273–279.

White L., Harvey P.D., Opler L., Lindenmayer J.P. (1997) Empirical assessment of the factorial structure of clinical symptoms in schizophrenia. A multisite, multimodel evaluation of the factorial structure of the Positive and Negative Syndrome Scale. The PANSS Study Group. *Psychopathology*, **30**(5), 263–274.

Wolthaus J.E., Dingemans P.M., Schene A.H., Linszen D.H., Knegtering H., Holthausen E.A. *et al.* (2000) Component structure of the Positive and Negative Syndrome Scale (PANSS) in patients with recent-onset schizophrenia and spectrum disorders. *Psychopharmacology (Berlin)*, **150**(4), 399–403.

Appendix: MPlus code for multilevel ordinal factor analysis model of White, *et al*. (1997).

Factor analytic model proposed by White *et al*. (1997)

```
Data: File is mplus.dat;
Variable: Names are rater pansp1 pansp2 pansp3 pansp4 pansp5 pansp6 pansp7 pansn1 pansn2 pansn3 pansn4 pansn5 pansn6 pansn7 pnsg1 pnsg2 pnsg3 pnsg4 pnsg5 pnsg6 pnsg7 pnsg8 pnsg9 pnsg10 pnsg11 pnsg12 pnsg13 pnsg14 pnsg15 pnsg16;
Cluster is rater;
Missing are all (99);
Categorical are pansn6 pansn1 pansn2 pansn3 pansn4 pnsg7 pnsg5 pnsg8 pnsg13 pansp1 pnsg9 pansp5 pansp3 pnsg1 pansp7 pnsg14 pansp4 pnsg2 pnsg4 pnsg3 pnsg6 pnsg11 pnsg15 pansn5 pansn7;
Usevariables are rater pansn6 pansn1 pansn2 pansn3 pansn4 pnsg7 pnsg5 pnsg8 pnsg13 pansp1 pnsg9 pansp5 pansp3 pnsg1 pansp7 pnsg14 pansp4 pnsg2 pnsg4 pnsg3 pnsg6 pnsg11 pnsg15 pansn5 pansn7;
Analysis: estimator = WLSMV; Type = twolevel;
Model:
%WITHIN%
Negative BY pansn6* pansn1 pansn2 pansn3 pansn4 pnsg7 pnsg5 pnsg8 pnsg13 pnsg14;
Negative@1;
Positive BY pansp1* pnsg9 pansp5 pansp3 pnsg1;
Positive@1;
Activat BY pansp7* pnsg14 pansp4 pnsg8 pansn3 pnsg4;
Activat@1;
Dysphoric BY pnsg2* pnsg4 pnsg3 pnsg6 pnsg1;
Dysphoric@1;
Autistic BY pnsg11* pnsg15 pansn5 pansn7 pnsg13 pansp3;
Autistic@1;
Output: Sampstat; Standardized; Residual; Tech1;
```