UCLA
UCLA Previously Published Works

Title
Measuring pathology using the PANSS across diagnoses: Inconsistency of the positive symptom domain across schizophrenia, schizoaffective, and bipolar disorder.

Permalink
https://escholarship.org/uc/item/1r27d940

Authors
Anderson, Ariana E
Mansolf, Maxwell
Reise, Steven P
et al.

Publication Date
2017-12-01

DOI
10.1016/j.psychres.2017.08.009

Peer reviewed
Measuring pathology using the PANSS across diagnoses: Inconsistency of the positive symptom domain across schizophrenia, schizoaffective, and bipolar disorder

Ariana E. Anderson\textsuperscript{a,b,*}, Maxwell Mansolf\textsuperscript{c}, Steven P. Reise\textsuperscript{c}, Adam Savitz\textsuperscript{d}, Giacomo Salvadore\textsuperscript{d}, Qingxin Li\textsuperscript{d}, Robert M. Bilder\textsuperscript{a,c}

\textsuperscript{a} University of California, Los Angeles, Department of Psychiatry and Biobehavioral Sciences, Los Angeles, CA, USA
\textsuperscript{b} University of California, Los Angeles, Department of Statistics, Los Angeles, CA, USA
\textsuperscript{c} University of California, Los Angeles, Department of Psychology, Los Angeles, CA, USA
\textsuperscript{d} Janssen Scientific Affairs, Titusville, NJ, USA

ARTICLE INFO

Keywords:
Factor analyses
Antipsychotic
Drug trials
Anxiety
Negative symptoms
Factor structure
Clinical trials

ABSTRACT

Although the Positive and Negative Syndrome Scale (PANSS) was developed for use in schizophrenia (SZ), antipsychotic drug trials use the PANSS to measure symptom change also for bipolar (BP) and schizoaffective (SA) disorder, extending beyond its original indications. If the dimensions measured by the PANSS are different across diagnoses, then the same score change for the same drug condition may have different meanings depending on which group is being studied. Here, we evaluated whether the factor structure in the PANSS was consistent across schizophrenia (n = 3647), bipolar disorder (n = 858), and schizoaffective disorder (n = 592). Along with congruency coefficients, Hancock's H, and Jaccard indices, we used target rotations and statistical tests of invariance based on confirmatory factor models. We found the five symptom dimensions measured by the 30-item PANSS did not generalize well to schizoaffective and bipolar disorders. A model based on an 18-item version of the PANSS generalized better across SZ and BP groups, but significant problems remained in generalizing some of the factors to the SA sample. Schizophrenia and bipolar disorder showed greater similarity in factor structure than did schizophrenia and schizoaffective disorder. The Anxiety/Depression factor was the most consistent across disorders, while the Positive factor was the least consistent.

1. Introduction

The Positive and Negative Syndrome Scale (PANSS), a clinician-reported outcome scale, is a commonly used instrument for measuring efficacy in anti-psychotic drug trials; the PANSS includes a total of 30 symptoms rated for severity on a 7-point scale (Kay et al., 1987). Although the PANSS was originally developed for use in schizophrenia (SZ), it is also used regularly for measuring treatment response in bipolar (BP) and schizoaffective (SA) disorder, often in trials examining the same medications that are used for treating schizophrenia (Berwaerts et al., 2011; Bossie et al., 2011; Canuso et al., 2009, 2010; Davidson et al., 2007; Gopal et al., 2010; Kramer et al., 2010, 2007; Meltzer et al., 2008; Tzimos et al., 2008; Vieta et al., 2010). Within schizophrenia, the PANSS is reported widely to comprise five dimensions, commonly described as the Positive, Negative, Excited, Disorganized, and Anxiety/Depression factors (Lindenmayer et al., 1995; Marder and Chouinard, 1997). The same factors observed in schizophrenia may not occur similarly in other syndromes, but the validity of this assumption has seldom been examined, despite the implications for evaluating efficacy. If the structure of symptoms differs across syndromes, then the same true effect size for a given intervention might yield different measured outcomes on the PANSS scales. Alternatively, different physiological effects might yield similar scale changes. It is important to be confident that we are measuring the same constructs, in the same way, across the different syndromes.

Earlier work on the PANSS factor structure compared raw scale scores between diagnostic groups on pre-specified dimensional scales, but did not determine if the same dimensions of psychopathology were being measured across syndromes. Daneluzzo and colleagues, for example, used an exploratory 3-factor model to compare BP and SZ patients, and reported higher Depression factor scores in BP but no difference in Negative factor scores (Daneluzzo et al., 2002). Difference in structure was not formally tested using statistical methods to detect measurement invariance across groups. Lindenmayer commented on
Results from target rotations are presented prior to performing conventional CFA models (measurement invariance analyses) because target rotations maximize the likelihood that the measured factor structure is similar across diagnoses compared to CFA, by allowing the unique factor structures within the test groups (schizoaffective disorder, bipolar disorder) to be rotated directly to the space of the schizophrenia group. This is complementary to traditional CFA approaches that measure whether the similarity between the observed and expected covariance matrices. In addition to the target rotations and the CFA-based measurement invariance tests presented here, in the Appendix we additionally perform CFA of extent models of the PANSS on the schizophrenia samples, further demonstrating that the five factor model using all, or nearly all, the items is difficult to reproduce on new SZ samples.

### 2.1. Data and demographics

The PANSS dataset consisted of baseline ratings from 16 Janssen Research & Development, LLC sponsored clinical trials, involving 5095 patients who had participated in the respective study for at least 21 days. All patients (SZ, SA, BP) were diagnosed according to DSM-IV criteria, and no patients met the diagnostic criteria for another Axis-1 diagnosis. The SZ patients in this sample were previously used in another study of the PANSS (Anderson et al., 2015). De-identified patient-level data with individual PANSS items were analyzed. Only baseline (pretreatment) measurements were used. Only 8 patients had some PANSS items missing, and median imputation was used to replace these values. The basic demographics and clinical characteristics of these patients are summarized in Table 1.

All studies were conducted in accordance to the latest version of the Declaration of Helsinki. After complete description of the study to the subjects, written informed consent was obtained. Different inclusion criteria were present depending on diagnoses, which we describe briefly below and in Supplementary Table 1. Centralized raters were not used for most studies. Study participants were enrolled in one of 12 trials (Supplemental Digital Content Table 1) who met all the following inclusion criteria: have a Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) diagnosis of schizophrenia (295.10, 295.20, 295.30, 295.60, or 295.90) for at least one year; experienced an acute episode of schizophrenia with a total PANSS score between 60 – 120 or 70 – 120 at screening, depending on the trial. Additional inclusion and exclusion criteria can be found at http://clinicaltrials.gov.

SCI-PANSS was not required for most of the studies. All raters in the study were required to have clinical experience and previous use of the PANSS. There was formal training either through in person investigator meetings or online or CD-ROM training for those who could not attend the investigator meeting. All raters had to pass a training video. All subjects were required to have informants. It was not a protocol requirement that informants be available at all interviews/assessments but the use of the informants was encouraged. It is not known how often informants were or were not used; this information was not captured in the data.

All except one SZ trial used 60–120 or 70–120 inclusion criteria, and SA studies had inclusion greater than 60, while for the bipolar studies, the PANSS was not used as inclusion/exclusion criteria and no severity range was specified. This explains in part why the total PANSS scores at baseline are higher in SZ and SA studies than BP studies. SA trials also required patients to have a score ≥ 4 on at least 2 of the following PANSS items: Hostility, Excitement, Tension, Uncooperativeness, and Poor Impulse Control at screening. Further trial details are published in the original articles (Berwaerts et al., 2011; Bossie et al., 2011; Canuso et al., 2009, 2010; Davidson et al., 2007; Gopal et al., 2010; Kramer et al., 2010, 2007; Meltzer et al., 2008; Tzimos et al., 2008; Vittinghoff et al., 2010). Analyses proposed here were repeated after applying the SA criterion to the BP and SZ populations to validate that the general conclusions were stable even when equivalent
<table>
<thead>
<tr>
<th>Internal ID</th>
<th>DX</th>
<th>NCT</th>
<th>N</th>
<th>Male (%)</th>
<th>PANSS (mean)</th>
<th>PANSS (SD)</th>
<th>Age in Years (Mean)</th>
<th>Age in Years (SD)</th>
<th>Inclusion</th>
</tr>
</thead>
</table>
| R076477-SCH-301| SZ | NCT00086320       | 164| 61.0%    | 92.10        | 11.3       | 39.78               | 9.8               | DSM-IV diagnosis of schizophrenia (295.10, 295.20, 295.30, 295.60, 295.90); Diagnosis of schizophrenia at least 1 year before screening; Experiencing an acute schizophrenic episode with a total PANSS score between 70 and 120, inclusive, both at screening and at baseline (the start of the run-in phase). DSM-IV diagnosis of schizophrenia (295.10, 295.20, 295.30, 295.60, 295.90) at least 1 year before screening. Total PANSS score at screening and baseline (day 0) between 70 and 120, inclusive. Subjects must have been diagnosed with schizophrenia according to DSM-IV (295.10, 295.20, 295.30, 295.60, 295.90) at least 1 year prior to screening. Subjects must be experiencing an acute episode, with a total PANSS score at screening between 70 and 120. Diagnosis of schizophrenia according to DSM-IV criteria (295.10, 295.20, 295.30, 295.60, 295.90) at least 1 year before screening; Experiencing an acute episode, with a total PANSS score at screening between 70 and 120. Experience an acute episode, with a total PANSS score at screening between 70 and 120; Diagnostic and Statistical Manual - Fourth Edition (DSM-IV) diagnosis of schizophrenia (paranoid, disorganized or undifferentiated type); score of $> = 4$ on at least two of a subset of selected PANSS items and a total score on these five items of $> = 17$; score of $> = 5$ on the CGI-S (clinical global impression - severity) Subjects diagnosed with schizophrenia according to DSM-IV [disorganized type (295.10), catatonic type (295.20), paranoid type (295.30), residual type (295.60), or undifferentiated type (295.90)] at least 1 year before screening. Total PANSS score must be between 70 and 120, inclusive, at screening, and 60 and 120, inclusive, at Day 1 (before start of double-blind study drug). Met diagnostic criteria for schizophrenia according to DSM-IV (disorganized type [295.10], catatonic type [295.20], paranoid type [295.30], residual type [295.60], or undifferentiated type [295.90]) for at least 1 year before screening. A total PANSS score between 60 and 120, inclusive, at screening and baseline. Met diagnostic criteria for schizophrenia according to DSM IV (disorganized type [295.10], catatonic type [295.20], paranoid type [295.30], residual type [295.60], or undifferentiated type [295.90]) for at least 1 year before screening. A total PANSS score at screening and at baseline of between 70 and 120, inclusive. Patients who meet diagnostic criteria for schizophrenia according to DSM IV for at least 1 year who meet PANSS score criteria and have body mass index (BMI) of $> = 15.0$ kg/m$^2$. PANSS total score at screening and baseline of 70-120, inclusive; Met diagnostic criteria for schizophrenia according to DSM-IV (disorganized type [295.10], catatonic type [295.20], paranoid type [295.30], residual type [295.60], or undifferentiated type [295.90]) for at least 1 year before screening. Prior medical records, written documentation or verbal information obtained from previous psychiatric providers obtained by the investigator must be consistent with the diagnosis of schizophrenia. A total PANSS score at screening of between 70 and 120, inclusive and at baseline of between 60 and 120, inclusive; Meets Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (DSM IV) criteria for Bipolar I Disorder, Most Recent Episode Manic or Mixed (with or without psychotic features); history of at least 1 previously documented manic or mixed episode requiring medical treatment within 3 years before the screening phase; total score of at least 20 on the Young Mania Rating Scale at screening and at baseline visit; Meets Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (DSM IV) criteria for Bipolar I Disorder, Most Recent Episode Manic or Mixed (with or without psychotic features); history of at least 1 previously documented manic or mixed episode requiring medical treatment within 3 years before the screening phase; total score of at least 20 on the YMRS at screening and at baseline visit; Meets Diagnostic and Statistical Manual of Mental Disorders - Fourth Edition (DSM-IV) criteria for Bipolar I Disorder, Most Recent Episode Manic or Mixed (with or without psychotic features); History of at least 1 previously documented manic or mixed episode requiring medical treatment within the past 3 years; Must have been taking lithium or valproate as part of treatment for Bipolar I Disorder for at least 2 weeks before screening, with drug levels at screening within therapeutic range; Total YMRS score of at least 20 at screening and at baseline (Day 1) Diagnostic and Statistical Manual - Fourth Edition (DSM-IV) diagnosis of schizoaffective disorder; A total Positive and Negative Symptoms of Schizophrenia (PANSS) score of $> = 60$; A score of $> = 16$ on Young Mania Rating Scale (YMRS) or a score of $> = 16$ on the Hamilton Depression Rating Scale (HAM-D 21 Diagnostic and Statistical Manual - Fourth Edition (DSM-IV) diagnosis of schizoaffective disorder; A total...

(continued on next page)
Inclusion criterion were applied.

### 2.2. Target rotations

In target rotations, one matrix is used as a target for rotation of the other matrices instead of applying other rotation criterion (e.g., Oblimin); this theoretically gives the other matrices the best chance of fitting the primary matrix’s factor structure, by aligning directly matrices to each other rather than indirectly using a third-target defined by some other criterion (e.g., Procrustes). We used the previously-identified 30-item factor structure within schizophrenia (Anderson et al., 2015) as the base for (partially specified) target rotations within schizoaffective and bipolar disorder. Parallel analysis (Horn, 1965) of the polychoric correlation matrix for the schizophrenia sample using all 30 PANSS items identified eight factors, with the final three being weak. In keeping with prior analyses of the PANSS (Lindenmayer et al., 1995), only five factors were used in subsequent analyses. This 30-item Oblimin-rotated model had the following symptom domains within schizophrenia:

- **Positive** = Delusions + Hallucinatory Behavior + Grandiosity + Suspiciousness Persecution + Unusual Thought Content + Preoccupation
- **Disorganized** = Stereotyped Thinking + Lack of Judgment and Insight + Conceptual Disorganization + Difficulty in Abstract Thinking + Mannerisms and Posturing + Poor Attention + Disturbance of Volition + Preoccupation + Disorientation
- **Excited** = Poor Impulse Control + Excitement + Hostility + Uncooperativeness
- **Anxiety** = Anxiety + Depression + Tension + Guilt Feelings + Somatic Concern

From the original Oblimin-rotated from loading matrix in schizophrenia (Anderson et al., 2015), a target matrix of equal dimension was constructed (Browne, 2001) to identify factor structures that were most likely to correspond across the three diagnostic samples. Using the notation of Browne (2001), all cells with values less than 0.2 in the rotated loading matrix were set to “0” in the target matrix, and all cells with values greater than 0.2 in the rotated loading matrix were set to “?” in the target matrix. Using this target matrix in target rotation, a rotated loading matrix is identified such that loadings with values of “0” in the target matrix are minimized, while loadings with values of “?” in the target matrix are freely estimated. The target matrix was then used in target rotations for the SA and BP samples, obtaining factor models for the structure seen in these disorders.

This process was performed separately for the 18-item and 30-item PANSS models. This 18-item model contained the following factors:

- **Negative** = Blunted Affect + Emotional Withdrawal + Poor Rapport + Passive Apathetic Social Withdrawal + Lack of Spontaneity and Flow of Conversation + Active Social Avoidance
- **Positive** = Delusions + Unusual Thought Content + Hallucinatory Behavior + Suspiciousness Persecution
- **Disorganized** = Stereotyped Thinking + Conceptual Disorganization + Poor Attention
- **Excited** = Poor Impulse Control + Hostility + Uncooperativeness
- **Anxiety** = Anxiety + Tension

### 2.3. Comparison of factor structure across diagnoses

We evaluated the consistency of the factor structure across diagnoses on the 18-item and the 30-item PANSS using: (1) Hancock’s H, which measures how well-defined a factor is within a sample, described
in more detail below (Hancock and Mueller, 2001, 2013). (2) Factor congruence, which measures pairwise consistency between schizophrenia and the other two disorders being evaluated (SA, BP). Factor congruence is a measure of factor similarity, and is mathematically defined as the cosine between two vectors of factor loadings, with values above 0.7 preferred (Ferguson and Cox, 1993). (3) The Jaccard Index (JI), which compares the consistency across all disorders being considered simultaneously (Milligan and Cooper, 1986). It is a ratio of the number of items in the intersection of the sets to the total number of items: \( JI(\text{SZ, BP, SA}) = \frac{\text{in common}}{\text{total} + \text{replicates} - \text{in common}} \). A Jaccard Index of 1 indicates the factor has identical symptom assignments to factors across diagnoses, while a low Jaccard index indicates inconsistency across diagnoses.

Hancock's H represents a measure of the degree to which individual dimensions have identical symptom assignments to factors across diagnoses, while a low Jaccard index indicates inconsistency across diagnoses, with 0 being an indication that two factors have no symptom in common, and (4) Measurement invariance which is based on traditional confirmatory factor analyses modeling, described in more detail below.

Hancock's H (Hancock and Mueller, 2001) were calculated for the target-rotated solutions. Hancock's H represents a measure of the degree to which individual dimensions have true individual differences on the factor (Hancock and Mueller, 2001). Possible values for Hancock's H range from 0 to 1, and values above 0.90 are preferred (Comrey and Lee, 1989). It is a ratio of the number of items in the intersection of the sets to the total number of items: \( H(\text{SZ, BP, SA}) = \frac{\text{in common}}{\text{total} + \text{replicates} - \text{in common}} \). A Jaccard Index of 1 indicates the factor has identical symptom assignments to factors across diagnoses, with 0 being an indication that two factors have no symptom in common, and (4) Measurement invariance which is based on traditional confirmatory factor analyses modeling, described in more detail below.

Hancock's H were calculated for the target-rotated solutions. Hancock's H represents a measure of the degree to which individual differences in factor score estimates reflect true individual differences on the factor (Hancock and Mueller, 2001). Possible values for Hancock's H range from 0 to 1, and values above 0.90 are preferred (Comrey and Lee, 1989). It measures how well-defined a factor is; a well-defined factor (with high H) would be expected to generalize, while factors that are not well-defined would not be expected to generalize. All analyses were performed using the free software R (R Core Development Team, 2012), using the psych (Revelle, 2011) and GPArotation (Bernaards and Jennrich, 2012) packages.

We assessed configural and weak measurement invariance (Meredith, 1993; Reise et al., 2000) using the model proposed in the exploratory factor analysis for the schizophrenia patient group in (Anderson et al., 2015). These models were fit using the DWLS estimation procedure and the “measurementInvariance” routine in R (R Core Development Team, 2012) from Lavaan (Rosseel, 2012). Configural invariance assesses whether the items have the same pattern of fixed and free loadings across factors. If configural invariance doesn't hold, then the PANSS is measuring different constructs across disorders, e.g., the Positive symptoms in schizophrenia are qualitatively different than Positive symptoms of schizoaffective disorder. If it does hold then similar, but not identical, latent variables are being measured across groups. Weak invariance tests whether the loadings are equal across groups, imposing a stronger constraint than configural invariance. If this is violated, then certain symptoms are more salient within factors than others. Model fit was measured using a chi-square value, the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA) (Ferguson and Cox, 1993), where invariance levels are compared using the change in CFI when imposing greater constraints on the underlying factor structure. CFI decreases as the fit decreases; a CFI value < 0.9 is considered a poor fit (Bentler, 1990). RMSEA increases as the fit decreases; a RMSEA value of zero indicates a perfect fit and RMSEA < 0.05 is considered a good fit. A change in CFI exceeding 0.02 suggests a violation of invariance (Cheung and Rensvold, 1999).

3. Results

The PANSS symptoms by dimension for each diagnosis are shown in Fig. 1 for the 30-item model. Using the Jaccard Index to compare three groups simultaneously, the factors in order from least to most consistent across diagnoses were Positive (JI = 0.30), Excited (JI = 0.38), Disorganized (JI = 0.58), Negative (JI = 0.64), and Anxiety/Depression (JI = 0.83). The Anxiety/Depression and Negative factors show relatively good consistency with no conspicuously variable items across diagnoses. For the Anxiety/Depression factor, the only inconsistent item across diagnoses was Excitment, which had a fairly strong loading (0.30) in the SZ group for the 30-item model. Similarly, two items (Guilt Feelings, Disorientation) loaded negatively on the Excited factor within the SA group, but did not load on this factor within the SZ or BP patient group. For the 18-item model, the Jaccard Indices were all higher than the 30-item model, with the exception of Disorganized; however, for the Disorganized domain some items very narrowly missed the 0.3 cutoff applied, so these items may be more consistent in new samples. The indices, in order from least to most consistent here: Disorganized (JI = 0.5), Positive (JI = 0.57), Negative (JI = 0.86), Excited (JI = 1) and Anxiety/Depression (JI = 1).

The target-rotated loadings for the 30-item PANSS for each diagnosis are shown in Table 2. For the five-factor solution on the SZ sample for the 30-item PANSS, the following fit statistics were calculated: the Root Mean Square of the Residuals (RMSR), the Tucker Lewis Index (TLI), and the RMSEA. The resulting values were RMSR = 0.04, TLI = 0.820, RMSEA = 0.067. The target-rotated loadings for the 18-item PANSS for each diagnosis are shown in Table 3. For the five-factor solution for the 18-item PANSS on the SZ sample, the following fit statistics were calculated: RMSR = 0.02, TLI = 0.898, RMSEA = 0.068. These solutions displayed numerous cross-loadings. Estimates of common variance (h²) reveal that between 9% and 72% of the variance in each variable is shared among the other variables in the 30-item PANSS, and between 32% and 76% of the variance in each variable is shared among the other variables in the 18-item PANSS. The average communality in the 30-item model was SZ = 0.44, SA = 0.42, BP = 0.56. The average communality for the 18-item model was SZ = 0.55, SA = 0.46, BP = 0.64. The large change in average communality between the 18-item and 30-item models suggest that some of the items in the full model don't share much in common with the rest of the items.

Very similar factor structures across the whole PANSS were identified in the schizophrenia and bipolar samples, but this structure did not replicate well in the schizoaffective sample. In the 30-item PANSS, Hancock's H generally did not meet the recommended value of 0.9 in the 18-item PANSS, while very similar factor structures were identified in the schizophrenia and schizoaffective samples. Factor congruency between the factors represented here are not well-determined.

In the 18-item PANSS, factor determinacy was highest in the schizophrenia and bipolar diagnosis groups; however, only two (Negative and Disorganized) factors in the bipolar sample demonstrated factor determinacy exceeding the recommended value of 0.9. Factor congruency between the factors estimated in the schizophrenia and bipolar samples was very high, with congruency coefficients ranging from 0.97 to 0.99. Factor congruency between the factors estimated in the schizophrenia and schizoaffective samples was also high, with the exception of factor F3 with estimated congruency of 0.87; other congruency values between these two samples ranged from 0.94 to 0.97. In short, while very similar factor structures were identified in the three samples, the factors represented here are not well-determined.

Table 4 shows the factor congruence across disorders for the 30-item model. Although the Negative factor was fairly consistent across disorders, the Positive and Excited factors were not. Factor congruency between the factors estimated in the schizophrenia and bipolar samples was high, with congruency coefficients ranging from 0.87 to 0.94. Factor congruency between the factors estimated in the schizophrenia and schizoaffective samples was modest; with the exception of factor Excited with estimated congruency of 0.53, other congruency values between these two samples ranged from 0.77 to 0.91. In the bipolar sample, and 0.66-0.88 in the schizoaffective sample. In the two samples in which the factor structure did replicate (schizophrenia and bipolar), Hancock’s H approached but rarely satisfied the 0.9 cutoff, indicating that while the factors identified by these structures are approaching determinacy, they are not well-determined, according to this criterion.

In the 18-item PANSS, factor determinacy was highest in the schizophrenia and bipolar diagnosis groups; however, only two (Negative and Disorganized) factors in the bipolar sample demonstrated factor determinacy exceeding the recommended value of 0.9. Factor congruency between the factors estimated in the schizophrenia and bipolar samples was very high, with congruency coefficients ranging from 0.97 to 0.99. Factor congruency between the factors estimated in the schizophrenia and schizoaffective samples was also high, with the exception of factor F3 with estimated congruency of 0.87; other congruency values between these two samples ranged from 0.94 to 0.97. In short, while very similar factor structures were identified in the three samples, the factors represented here are not well-determined.

Table 4 shows the factor congruence across disorders for the 30-item model. Although the Negative factor was fairly consistent across disorders, the Positive and Excited factors were not. Factor congruency between the factors estimated in the schizophrenia and bipolar samples was high, with congruency coefficients ranging from 0.87 to 0.94. Factor congruency between the factors estimated in the schizophrenia and schizoaffective samples was modest; with the exception of factor Excited with estimated congruency of 0.53, other congruency values between these two samples ranged from 0.77 to 0.91.

Table 5 shows the factor congruence across disorders for the 18-item model. These reduced factors were all strongly correlated with each other across diagnoses, ranging from 0.79 to 1. This lowest correlation values with SZ were once again in the Positive domain with SA...
The remaining correlation values, all were above 0.93. For measurement invariance tests, the fit indices capture the average fit of models across diagnoses—models derived originally in SZ are fit to SZ, SA, and BP through confirmatory models. For configural invariance of the 30-item SZ-derived model, the model had a cross-diagnoses CFI of 0.663, RMSEA of 0.070, a chi-square value of 10199 (df = 1014). For weak invariance (equal loadings), the model had CFI of 0.641, RMSEA of 0.071, and chi-square value of 13162 (df = 1064). For model comparison, the change in CFI of 0.022 and the change in chi-square both suggest that weak invariance is violated across diagnoses. 

Fig. 1. Factor domains for the 30-item PANSS by diagnoses, using the 0.3 threshold. Using the Jaccard Index, the factors in order from least to most consistent across diagnoses were Positive (JI = 0.25), Disorganized (JI = 0.31), Negative (JI = 0.5), Excited (JI = 0.67), and Anxiety/Depression (JI = 0.83).
<table>
<thead>
<tr>
<th>Negative</th>
<th>Positive</th>
<th>Excited</th>
<th>Disorganized</th>
<th>h²</th>
</tr>
</thead>
<tbody>
<tr>
<td>SZ</td>
<td>SA</td>
<td>BP</td>
<td>SZ</td>
<td>SA</td>
</tr>
<tr>
<td>Blunted Affect</td>
<td>-0.14</td>
<td>0.27</td>
<td>-0.11</td>
<td>0.51</td>
</tr>
<tr>
<td>Emotional Withdrawal</td>
<td>0.80</td>
<td>0.77</td>
<td>0.77</td>
<td>-0.06</td>
</tr>
<tr>
<td>Lack of Spontaneity and Flow of Conversation</td>
<td>0.55</td>
<td>0.70</td>
<td>0.88</td>
<td>-0.11</td>
</tr>
<tr>
<td>poor Impulse Control</td>
<td>0.02</td>
<td>0.14</td>
<td>0.15</td>
<td>-0.08</td>
</tr>
<tr>
<td>Disturbance of Volition</td>
<td>0.30</td>
<td>0.36</td>
<td>-</td>
<td>0.17</td>
</tr>
<tr>
<td>Preoccupation</td>
<td>0.34</td>
<td>0.30</td>
<td>0.25</td>
<td>-0.07</td>
</tr>
<tr>
<td>Disorientation</td>
<td>0.03</td>
<td>0.09</td>
<td>0.14</td>
<td>-0.02</td>
</tr>
<tr>
<td>Excitement</td>
<td>-0.23</td>
<td>-0.29</td>
<td>-</td>
<td>0.16</td>
</tr>
<tr>
<td>Hostility</td>
<td>0.30</td>
<td>0.31</td>
<td>0.27</td>
<td>-0.04</td>
</tr>
<tr>
<td>Anxiety</td>
<td>0.27</td>
<td>0.33</td>
<td>0.24</td>
<td>-0.04</td>
</tr>
<tr>
<td>Guilt Feelings</td>
<td>0.10</td>
<td>0.13</td>
<td>0.08</td>
<td>-0.04</td>
</tr>
<tr>
<td>Tension</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>-0.03</td>
</tr>
<tr>
<td>Depression</td>
<td>0.03</td>
<td>0.06</td>
<td>0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td>Hancock's H</td>
<td>0.87</td>
<td>0.88</td>
<td>0.91</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Table 2: Factor by diagnosis for 30-item PANSS.
For configural invariance of the 18-item SZ-derived model, the model had a cross-diagnoses CFI of 0.723, RMSEA of 0.081, a chi-square value of 3812.3 (df = 375). For weak invariance (equal loadings), the model had CFI of 0.714, RMSEA of 0.079, and chi-square value of 4448.9 (df = 401). For model comparison, the change in CFI of 0.010 indicates that reducing the current diagnostic criteria. We found, however, that reducing the number of factors in the BP group from five to four did not eliminate the Disorganized factor, suggesting that the Disorganized dimension is distinct yet may be weak in BP. In the Disorganized factor, seven of the thirteen items showed consistent loadings across diagnoses, suggesting diagnostic groups.

### 4. Discussion

We applied multiple statistical strategies in large groups of patients with diagnoses including schizophrenia, schizoaffective, and bipolar disorder, to assess the dimensions of psychopathology measured by the PANSS, one of psychiatry's most widely used symptom rating scales. The 30-item factor structure varied across diagnostic groups even with target factor rotations, but the 18-item factor structure showed better concordance at least between schizophrenia and bipolar disorder. When running an even more parsimonious model of the PANSS containing 16 items and 4 factors (removing Anxiety and Tension), the consistency of the reduced PANSS across diagnoses was similarly demonstrated. Surprisingly, larger differences were found between SZ and SA than between SZ and BP; SA was less consistent than the other disorders and had lower factor determinacy, and may be more heterogeneous than the other disorders since it includes both depressive and manic symptoms.

The Positive factor was the most inconsistent across schizophrenia, schizoaffective, and bipolar disorder diagnoses. Factor models of the Positive domain need to include different symptoms depending on the diagnoses being assessed; Positive symptoms in SA should contain measures of Hostility, but including these symptoms would be inappropriate for assessing Positive symptoms of BP, since it would alter reliability by including loadings which have limited explanatory power (negligible variance). For the Positive factor, only three items (Hallucinatory Behavior, Delusions, Suspiciousness/Persecution) had consistent loadings across diagnoses. The Positive symptoms may be less stable across diagnoses depending on how the patient populations are selected, consistent with the low Hancock’s H value for this sample.

The Excited factor was similarly inconsistent across diagnosis. Only three of nine symptoms were consistently present across disorders (Hostility, Uncooperativeness, Poor Impulse Control). Schizoaffective disorder was unique in that it included both depressive and manic symptoms. For the Positive factor, only three items (Hallucinatory Behavior, Delusions, Suspiciousness/Persecution) had consistent loadings across diagnoses. The Positive symptoms may be less stable across diagnoses depending on how the patient populations are selected, consistent with the low Hancock’s H value for this sample.

The Excited factor was similarly inconsistent across diagnosis. Only three of nine symptoms were consistently present across disorders (Hostility, Uncooperativeness, Poor Impulse Control). Schizoaffective disorder was unique in that it included both depressive and manic symptoms. For the Positive factor, only three items (Hallucinatory Behavior, Delusions, Suspiciousness/Persecution) had consistent loadings across diagnoses. The Positive symptoms may be less stable across diagnoses depending on how the patient populations are selected, consistent with the low Hancock’s H value for this sample.

The Excited factor was similarly inconsistent across diagnosis. Only three of nine symptoms were consistently present across disorders (Hostility, Uncooperativeness, Poor Impulse Control). Schizoaffective disorder was unique in that it included both depressive and manic symptoms. For the Positive factor, only three items (Hallucinatory Behavior, Delusions, Suspiciousness/Persecution) had consistent loadings across diagnoses. The Positive symptoms may be less stable across diagnoses depending on how the patient populations are selected, consistent with the low Hancock’s H value for this sample.

The Disorganized factor also exhibited moderate variability. It was previously noted (Toomey et al., 1998) that the BP group lacked a Disorganized factor altogether, which may be seen as consistent with current diagnostic criteria. We found, however, that reducing the number of factors in the BP group from five to four did not eliminate the Disorganized factor, suggesting that the Disorganized dimension is distinct yet may be weak in BP. In the Disorganized factor, seven of the thirteen items showed consistent loadings across diagnoses, suggesting
that disorganization may be a shared phenotype across disorders that is even more common than the Positive or Excited symptoms.

There are several limitations to this work: the factor structure depends on the patient sample which differed across studies even holding constant the PANSS inclusion criteria; for example, mania when present within bipolar disorder would cause this factor structure to differ from schizophrenia and schizoaffective disorder. Different patient subtypes will yield qualitatively different results, even when inclusion criteria stipulating the total severity of psychopathology (total PANSS) are held constant. These results then are heavily dependent both on inclusion criterion and the natural variability present within these disorders.

The inconsistency of factor structure in the original 30-item PANSS indicates that it should be used cautiously across diagnostic groups, especially when comparing total scores or differential total scores across patient groups. Pooling patients across disorders in a single analysis might blur effect sizes by introducing systematic measurement variance; when factors derived in SZ are used to describe other groups, the factor scores will include contributions from PANSS items that are unrelated to the underlying factor, introducing noise into estimates of the patients’ true scores on that dimension. Assuming a common factor model across diagnoses invariably also omits symptoms that are important to a given dimension, decreasing signal and reducing statistical power. Together, these measurement problems result in lower-quality and less precise measurements of the specific symptom domains of interest.

Variation in PANSS factor structure across diagnoses does not imply that the PANSS instrument cannot be used to measure disorders other than schizophrenia; rather, it means that the PANSS instrument measures symptom constructs differentially and uniquely within each disorder. The 18-item PANSS provided a more stable framework for cross-diagnostic comparison. In future work, we propose using alternate psychometric strategies (e.g., bifactor models) to determine whether incorporating a “general symptom” dimension will better explain shared symptom structure, while also permitting specification of individual psychopathological dimensions that may be unique to different groups of patients.

Acknowledgements

The authors thank the study volunteers for participating in the research studies and to the clinicians and support staff for enabling patient recruitment and clinical data collection. The authors also thank Monique Franc, Anthony Santos, Nicole Bottel, William Cafferty and other staff in the Neuroscience Biomarkers of Janssen Research & Development for operational support including anonymization of clinical data.

Table 4

<table>
<thead>
<tr>
<th>Negative</th>
<th>Positive</th>
<th>Disorganized</th>
<th>Excited</th>
<th>Anxiety/Depression</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PANSS Factor</strong></td>
<td><strong>SZ</strong></td>
<td><strong>SA</strong></td>
<td><strong>BP</strong></td>
<td><strong>SZ</strong></td>
</tr>
<tr>
<td>Negative</td>
<td>1.00</td>
<td>0.96</td>
<td>0.94</td>
<td>0.40</td>
</tr>
<tr>
<td>Positive</td>
<td>-0.22</td>
<td>-0.31</td>
<td>-0.34</td>
<td>1.00</td>
</tr>
<tr>
<td>Negative</td>
<td>0.96</td>
<td>1.00</td>
<td>0.96</td>
<td>0.31</td>
</tr>
<tr>
<td>Positive</td>
<td>0.94</td>
<td>0.96</td>
<td>1.00</td>
<td>-0.34</td>
</tr>
</tbody>
</table>

Table 5

<table>
<thead>
<tr>
<th>Negative</th>
<th>Positive</th>
<th>Disorganized</th>
<th>Excited</th>
<th>Anxiety/Depression</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PANSS Factor</strong></td>
<td><strong>SZ</strong></td>
<td><strong>SA</strong></td>
<td><strong>BP</strong></td>
<td><strong>SZ</strong></td>
</tr>
<tr>
<td>Negative</td>
<td>1.00</td>
<td>0.96</td>
<td>0.94</td>
<td>0.35</td>
</tr>
<tr>
<td>Positive</td>
<td>-0.40</td>
<td>-0.43</td>
<td>-0.46</td>
<td>1.00</td>
</tr>
<tr>
<td>Negative</td>
<td>0.96</td>
<td>0.97</td>
<td>0.97</td>
<td>0.43</td>
</tr>
<tr>
<td>Positive</td>
<td>0.94</td>
<td>0.97</td>
<td>1.00</td>
<td>-0.46</td>
</tr>
<tr>
<td>Disorganized</td>
<td>0.35</td>
<td>-0.35</td>
<td>-0.44</td>
<td>0.83</td>
</tr>
<tr>
<td>Excited</td>
<td>-0.44</td>
<td>-0.44</td>
<td>-0.47</td>
<td>0.98</td>
</tr>
<tr>
<td>Anxiety/Depression</td>
<td>-0.08</td>
<td>-0.38</td>
<td>-0.00</td>
<td>-0.11</td>
</tr>
</tbody>
</table>

18-item factor congruence.
Interface from BWF.

References


