Title

Sparse factors for the positive and negative syndrome scale: Which symptoms and stage of illness?

Journal

Psychiatry Research, 225(3)

ISSN

0165-1781

Authors

Anderson, Ariana
Wilcox, Marsha
Savitz, Adam
et al.

Publication Date

2015-02-01

DOI

10.1016/j.psychres.2014.12.025

Peer reviewed

HHS Public Access

Author manuscript

Psychiatry Res. Author manuscript; available in PMC 2016 February 28.

Published in final edited form as:

Psychiatry Res. 2015 February 28; 225(3): 283–290. doi:10.1016/j.psychres.2014.12.025.

Sparse factors for the positive and negative syndrome scale: Which symptoms and stage of illness?

Ariana Anderson[a,*], **Marsha Wilcox**[b], **Adam Savitz**[b], **Hearee Chung**[c], **Qingqin Li**[b], **Giacomo Salvadore**[b], **Dai Wang**[b], **Isaac Nuamah**[b], **Steven P. Riese**[d], and **Robert M. Bilder**[a]

[a]Department of Psychiatry and Biobehavioral Sciences, University of California, 760 Westwood Plaza, C8-739 Semel Institute, Los Angeles, CA, USA

[b]Janssen Research and Development, Titusville, NJ, USA

[c]Janssen Scientific Affairs, Titusville, NJ, USA

[d]Department of Psychology, University of California, Los Angeles, CA, USA

Abstract

The Positive and Negative Syndrome Scale (PANSS) is frequently described with five latent factors, yet published factor models consistently fail to replicate across samples and related disorders. We hypothesize that (1) a subset of the PANSS, instead of the entire PANSS scale, would produce the most replicable five-factor models across samples, and that (2) the PANSS factor structure may be different depending on the treatment phase, influenced by the responsiveness of the positive symptoms to treatment. Using exploratory factor analysis, confirmatory factor analysis and cross validation on baseline and post-treatment observations from 3647 schizophrenia patients, we show that five-factor models fit best across samples when substantial subsets of the PANSS items are removed. The optimal model at baseline (five factors) omits 12 items: Motor Retardation, Grandiosity, Somatic Concern, Lack of Judgment and Insight, Difficulty in Abstract Thinking, Mannerisms and Posturing, Disturbance of Volition, Preoccupation, Disorientation, Excitement, Guilt Feelings and Depression. The PANSS factor models fit differently before and after patients have been treated. Patients with larger treatment response in positive symptoms have larger variations in factor structure across treatment stage than the less responsive patients. Negative symptom scores better predict the positive symptoms scores after treatment than before treatment. We conclude that sparse factor models replicate better on new samples, and the underlying disease structure of Schizophrenia changes upon treatment.

Keywords

PANSS; Confirmatory factor analysis; Exploratory factor analysis; Schizophrenia; RDoC; Dimensional Measures

*Correspondence to: Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, 760 Westwood Plaza, C8-739 Semel Institute, Los Angeles, CA 90095, Tel.: +1 310 254 5680, fax: +1 310 206 1866.

Appendix A. Supporting information: Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.psychres.2014.12.025.

1. Introduction

The Positive and Negative Syndrome Scale (PANSS) is a well-established scale for evaluating symptom severity in Schizophrenia (Kay et al., 1987), measuring 30 separate items such as “Hallucinatory Behavior” and “Blunted Affect.” Five latent dimensions or constructs, commonly referred to as factors, are assumed to underlie these 30 symptoms (White et al., 1997; Meyer, 2003; Van den Oord et al., 2006; Aggarwal et al., 2011), yet these proposed dimensions lack consistency across studies (Lehoux et al., 2009).

Many published factor models of Schizophrenia, empirically produced by exploratory factor analysis (EFA), subsequently fail to replicate across samples. Using confirmatory factor analysis (CFA), a recent study examined 25 published 5-factor models on a new dataset containing nearly 6000 patients from clinical trials (van der Gaag et al., 2006); none of the models fit the data well. A more recent study examined 29 published models with a new sample; again, none fit the data well (Wallwork et al., 2012). Failure of replication is an all-too-common criticism of factor structures proposed for the PANSS.

It has been hoped that the factor structure of the PANSS or other rating scales will reflect latent dimensions of Schizophrenia and/or other mental disorders. Establishing the consistency of these dimensions is essential for the Research Domain Criteria project (RDoC), which “to develop a research classification system for mental disorders based upon dimensions of neurobiology and observable behavior” (Cuthbert and Insel, 2013). These factors of the PANSS could directly provide dimensional measures of psychotic illness (Cuthbert and Insel, 2013) by using the resulting factor scores on each dimension.

There are several reasons why the factors found in one study could fail to replicate in another. Factorial change could suggest instability in the consistency and accuracy of the actual measurement (Meredith, 1993; Vandenberg and Lance, 2000), a violation of methodological assumptions such as ordinality (Kelley et al., 2013) or differences in the subpopulations being studied (Khan et al., 2014). Factorial invariance is analogous to measuring weight using a different scale across subjects, where true differences among patients and treatments may be either blurred or artificially introduced by the variability among the measurement instruments used. Because of this, it is important to understand why the proposed factor models of the PANSS often fail to replicate across studies, and why the hypothesized dimensions of Schizophrenia, as defined by the PANSS, fail to reappear when using different samples.

The pyramidical model of the PANSS retained 25 items and four factors from an EFA analysis, omitting factors with smaller eigenvalues “since these latter components are likely to be describing error variances or factors of minor influence (Kay and Sevy, 1990).” In related work using CFA, we showed that none of five published factor models fit data from 3647 Schizophrenia patients enrolled in randomized clinical trials (Anderson et al., under review). The best performing model in our CFA analysis, the Pentagonal model, retained 25 of the 30 PANSS items (White et al., 1997), while the Marder model (Marder et al., 1997) containing all 30 items performed poorly, based on conventional goodness-of-fit statistics (root mean square error of approximation [RMSEA], and confirmatory fit index [CFI]).

Hayashi (Hayashi et al., 2002) reported that using fewer than half of the PANSS items produced the most resilient models across gender. These findings together suggest that models containing fewer items (but not necessarily fewer factors) may be more robust on new samples than models using all 30 items.

We hypothesize that PANSS factor models may replicate better on new samples when they include fewer items. Our secondary hypothesis is that the PANSS the factor structure might change before and after treatment, partially due to the prominence of positive symptoms prior to treatment. A change in the factor structure is additionally supported by the recent finding that the PANSS item relationships differ by illness phase (Khan et al., 2014). Collectively, this paper assesses whether the failure of PANSS models to replicate across samples is due to low-loading individual items, and whether the symptom structure measured by the PANSS might change in response to treatment because of the disproportionate response of the positive symptom domain to interventions.

2. Methods

We used clinical trial data from 3647 unique schizophrenia patients who participated in a medication treatment trial for at least 21 days, gathered from a total of 11 studies. Written informed consent for all patients was obtained after the study procedure was fully explained. Together, these studies examined six different treatments and included 36 countries. The demographic summary information for each study is presented in Table 1, detailed further in (Anderson et al., under review). A total of 10 PANSS administrations (out of 109,410) were missing from eight subjects and were imputed by using the overall median of other PANSS items within that patient. Baseline and post-treatment data were assessed separately. PANSS items with a score of “seven” were rare, and were recorded as “six” to increase stability of the subsequent analyses. We performed this analysis within R (R Development Core Team, 2013) using the packages pysch (Revelle, 2011), semTools (Pornprasertmanit et al., 2013) and lavaan (Rosseel, 2012).

2.1. The Replicability of full and restricted PANSS factor models across samples

We assessed the comparative validity of full and restricted PANSS models using two different approaches. Firstly, we used cross-validated EFA/CFA models to assess whether models which include fewer PANSS items replicate better on new samples, varying not just the sampling partition, but also the rotation method and thresholding procedure used to decide which PANSS items were retained in the factor model. Secondly, we compared the fit of two 30-item PANSS factor models and two restricted PANSS models containing 18 and 20 items, using the Bayesian Information Criterion along with the CFI, RMSEA, and the SRMR. Through this, we determined whether the model fit depends on the selection of items used, and whether subsets of PANSS items might replicate better on new samples.

We first tested whether sparser factor models fit better on new samples by cross-validating EFA derived factor models with a separate CFA, using the CFI of each parameter setting to measure how well the proposed model fit the new sample. We evaluated whether excluding certain PANSS items may increase the stability of the traditional PANSS 5-factor models across data samples, by varying the thresholds used for item inclusion between 0.2 and 0.65

by 0.05. The thresholds act as a gateway for items being included in a model (e.g., the item Hallucinations would be assigned to the Positive factor when the loading for Hallucination on the Positive factor was greater than the threshold). High thresholds can prevent items from contributing to more than one construct, resulting in sparse factor models. Moderate thresholds usually include all items but may result in items loading on only one factor. Low thresholds may include all items and allow items to influence more than one factor (cross-loading), thereby revealing inter-factor correlation. By varying thresholds, we thus additionally tested whether models that allow cross-loadings are more resilient on new data samples.

Statistical models usually have superior fit on the dataset to which they were trained, allowing sampling variability to be the cause of replication failure (Efron and Efron, 1982). Cross-validation is a common statistical technique used to estimate prediction error in many other fields, although it has only rarely been used in psychometric analyses of the PANSS (van der Gaag et al., 2006). In 10-fold cross-validation, the data are partitioned into 10 folds, and models are trained using 90% of the data and tested (validated) on the remaining 10%. The average fit statistic over the 10 partitions is the "*generalization error when the method is applied to an independent test sample from the joint distribution of X and Y*" (Hastie et al., 2009).

Applying 10-fold cross-validation here, we partitioned the schizophrenia baseline patients into 10 subgroups randomly. For each partition we performed an EFA on 90% of the data using polychoric correlations to account for the ordinal nature of the PANSS and tested it in a CFA on the remaining 10% of observations. We cycled through these partitions 10 times with different training and validation datasets each time, where a total of 100% of the data was used as a "testing" set. The average CFI fit over all partitions provides an estimate of how well the proposed factor model with a given threshold would perform on a new dataset, for a specific rotation method used (under the same conditions as the model used). This process was repeated separately using the post-treatment observations.

We used four different factor rotation methods to examine the sensitivity of model fit to rotation: orthogonal (Varimax) and oblique (Promax, Oblimin and Non-negative Matrix Factorization). The factor rotations rearrange the original mathematically-derived loadings to make the resulting patterns more intuitive; for example, the Varimax rotation enforces that each factor has a small number of large loadings and a large number of small loadings, basically ensuring that for each factor, a limited number of items will be associated with it. We additionally analyzed the effect of estimation procedures (robust vs. non-robust) for the Promax rotation using robust diagonally weighted least squares (DWLS) (Yang-Wallentin et al., 2010).

With the 10 partitions for creating and testing models, four rotation methods, 10 thresholds, and two treatment stages being evaluated, a total of 800 models were created and validated on separate data sets, to evaluate the underlying reason for the frequent failure of PANSS models to replicate. The average CFI for each partition is shown in Fig. 1. Through this, we observed whether the PANSS factor models replicated better on new data using sparser

models (i.e., with higher thresholding of the loading matrix), independent of the effects of the rotation methods and sampling variability.

We next tested whether removing PANSS items produced better fitting factor models by comparing the fit of a 30-item model, derived using the cross-validated factor loadings on the baseline data ("full model"), with the fit of an 18-item model created with a higher threshold on the same cross-validated factor loadings. Each model was tested separately on the pre-treatment and post-treatment data using a CFA fit using DWLS to estimate the model parameters (Yang-Wallentin et al., 2010), using the full weight matrix to compute robust standard errors, and a mean- and variance-adjusted test statistic. For each CFA we extracted the chi-square value, RMSEA, SRMR (Standardized Root Mean Residual), CFI, and Bayesian Information Criterion (BIC). The BIC calculation was computed separately from the other fit statistics using the maximum likelihood estimation. We formally compared the fit of competing models using the BIC: a 30-item factor model proposed by this dataset with a restricted 18-item factor model, where a lower BIC indicates either better fit, fewer explanatory variables, or both, and supports sparse models. A BIC difference greater than 10 indicates strong evidence for the competing model (Kass and Raftery, 1995). For CFI, a value of >0.9 is considered a good fit (Bentler, 1990). For RMSEA, a value of zero indicates a perfect fit with RMSEA <0.05 being considered a good fit, and RMSEA increases as the fit decreases (Browne et al., 1993; Chen et al., 2008). These comparisons were performed on both the baseline and post-treatment data, with the results of these tests provided in Table 4.

2.2. Change in PANSS factor structure across treatment stage

We assessed whether the fit of the factor models differed by treatment stage for the 30-item full model proposed by this data, the 30-item Marder model (Marder et al., 1997), and the restricted models proposed separately for the baseline and post-treatment data from the cross-validated analysis. Formally, we assumed the null hypothesis that the RMSEA obtained from the CFA was similar before and after treatment, holding constant the model being evaluated. To test equality of two parameters when the standard errors are approximately equal at the 5% significance level, approximately 83% confidence intervals of the parameters can be compared for overlap (Payton et al., 2003). We computed 90% confidence intervals, which further reduces the Type 1 error below the 5% significance level. These confidence intervals were compared within factor model, across treatment stages; if for a given model the pre-treatment and post-treatment RMSEA confidence intervals obtained through a CFA did not overlap, we then rejected the null hypothesis that the RMSEA for that model was similar across the baseline and post-treatment observations, at the 5% significance level.

The resulting confidence intervals for the models, before and after treatment, are provided in Table 4. We also measured the fit of the restricted baseline, restricted post-treatment, full, and Marder models (Marder et al., 1997) across treatment stage using the CFI, SRMR and the BIC. These results are supplied in Table 4. For the baseline and post-treatment data, the squared multiple correlations are additionally provided as Table 5; the squared multiple

correlation of a variable with the remaining variables in a matrix is sometimes used as initial estimates of the communality of a variable.

As a technical note, fit statistics derived in a CFA do not test whether, for a given threshold, the same items fall into the same factors. Instead, the fit statistics measure differences in the covariance structure of the observed symptoms compared to the covariance structure suggested by the model. These fit statistics are then compared within a single model, across treatment stages.

We hypothesized that the change in factor structure was caused by the responsiveness of positive symptoms to treatment. To assess this, we segmented the patients into two groups using median-split based on the patient's total change in positive symptoms, and performed a CFA using the full model (proposed by the entire dataset) on each of the subgroups: high-response pre-treatment, high-response post-treatment, low-response pre-treatment, low-response post-treatment. A larger difference in fit across treatment stage for the patients who also demonstrate large changes in their positive symptom subscale scores would support this hypothesis.

Finally, we compared the relationship of the positive and negative symptoms subscales before and after treatment, using a general linear mixed effects model fit using restricted maximum likelihood: we predicted the total positive symptom subscale using the negative symptom subscale score, the treatment phase, and the interaction between treatment phase and negative symptom subscale scores as fixed effects, modeling the patient as the random effect. The interaction effect, between the negative symptoms and the treatment stage in predicting the positive symptoms, tests whether the relationship of the positive and negative items (subscales) differs based upon treatment. A significant slope is directly related to the covariance and the correlation between the negative and positive symptoms (using $\beta = r(\sigma_y/\sigma_x)$). Change in the regression slope between positive and negative symptoms based on treatment phase is additional evidence for the change in factor structure depending on treatment phase.

3. Results

The PANSS factor models with fewer symptoms had the highest CFI fit on new data samples. The strongest PANSS models were those with high thresholds, which excluded 40% of the total PANSS items, shown in Fig. 1. Low-threshold models also performed well, which included all PANSS items with cross-loadings when the loadings were above the threshold. Models which were created using moderate thresholds typically contained all symptoms but no cross-loading items, and performed the worst. This suggests that the most reproducible five-factor models are those that remove PANSS items, but alternatively retaining all PANSS items while allowing cross-loadings provides more resilient models than those models that allow symptoms to map to only one factor.

Thresholding models at 0.55 led to roughly 12 PANSS items being removed in the baseline data (depending upon cross-validation iteration and rotation criterion used). Increasing this to 0.65 led to roughly 10 retained PANSS items (of the original 30) and provided optimum fit, yet frequently did not allow model convergence since some factors (especially

disorganized) were removed entirely as they did not contain any items. Rotation method did not significantly influence model resiliency and retained similar variance levels, although NMF performed substantially worse than both oblique and orthogonal rotation methods. Secondary estimation using the DWLS (robust) optimization technique yielded a statistical tie for the thresholds of 0.55 and 0.50, with corresponding average CFI values of 0.8574 and 0.8592. Based on this, we recommend the EFA threshold of 0.55, for all rotation methods and optimization procedures assessed here.

The optimal baseline reduced-PANSS factor model (Oblimin rotation) with the full loadings (averaged across 10 data folds) is provided in Table 2. We defined the optimal model as that which retained highest fit while still converging across all 10-samples. This model does not contain cross-loading items because of the high-threshold used.

- 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

The eliminated items at baseline were Motor Retardation, Grandiosity, Somatic Concern, Lack of Judgment and Insight, Difficulty in Abstract Thinking, Mannerisms and Posturing, Disturbance of Volition, Preoccupation, Disorientation, Excitement, Guilt Feelings and Depression.

The post-treatment factor model contained 20 PANSS items (Oblimin rotation) with a threshold of.55, which is presented as follows. The full loadings (averaged across 10 data folds) are provided in Table 3.

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

The eliminated items at post-treatment were Grandiosity, Stereotyped Thinking, Somatic Concern, Lack of Judgment and Insight, Difficulty in Abstract Thinking, Mannerisms and Posturing, Disturbance of Volition, Preoccupation, Disorientation, Excitement, Hostility.

The “Full Model” which assigned every PANSS item to exactly one domain based on the maximal average cross-validated pre-treatment loading matrix is

- Negative=Blunted Affect+Emotional Withdrawal+Poor Rapport+Passive Apathetic Social Withdrawal+Lack of Spontaneity and Flow of Conversation+Motor Retardation+Active Social Avoidance
- Positive=Delusions+Unusual Thought Content+Hallucinatory Behavior +Suspiciousness Persecution+Grandiosity
- Disorganized=Conceptual Disorganization+Poor Attention+Mannerisms and Posturing+Difficulty in Abstract Thinking+Disturbance of Volition+Preoccupation +Disorientation+Stereotyped Thinking+Lack of Judgment and Insight
- Excited=Excitement+Poor Impulse Control+Hostility+ Uncooperativeness
- Anxiety=Anxiety+Tension+Depression+Guilt Feelings+Somatic Concern

Each model was tested separately on the pre-treatment and post-treatment data in a CFA, as shown in Table 4. For all models, the BIC was markedly reduced for the models with fewer items (BIC difference exceeding 10), providing strong evidence for the simplified factor models. Other fit indices such as the CFI, RMSEA, and SRMR also showed improved fit for more parsimonious models.

To evaluate whether the factor models differed by treatment stage, we computed 90% confidence intervals of the population RMSEA for the CFA using data from before and after treatment. Non-overlapping 90% confidence intervals for a given model implied that the model fits were significantly different before and after treatment, holding constant the model. The confidence intervals (Table 4) do not overlap for the full models, so we reject the null hypothesis and conclude that the factor structure differs across treatment stages ($p < 0.05$). The confidence intervals for the reduced models did not overlap for the non-robust estimation method but did overlap for the robust estimation method; based on this inconsistency we do not reject the null hypothesis for the restricted models.

We computed the squared multiple correlation to assess how the items varied with other items, before and after treatment, which are used as communality estimates. After treatment, symptoms were more strongly correlated with the remaining symptoms for all symptom domains, as shown in Table 5. This suggests that the treatment decreased the variability seen across symptoms, or could be a “floor” effect where the variance decreased since patients total score was at the lower end of the spectrum.

We hypothesized that the responsiveness of the positive symptoms to treatment may be responsible for the variability in factor structure across treatment stage, and tested this by performing a CFA using the full model in both high and low-responders (split by median positive symptom change). The CFI of the fits for each partition are shown in Table 6. Patients who had higher changes in positive symptoms had a change of .13 across treatment stage, compared to a change of .043 in patients who had lower positive changes. This suggests that patients who show high positive symptom treatment response also show more variability in their factor structure across treatment.

Finally, we tested whether the correlation of the negative and positive symptoms changed after treatment, using a general linear mixed-effects model including patient ID as the random effect (Table 7). The significant interaction effect between the negative symptoms and the treatment stage indicates that the relationship between these domains is significantly different before and after treatment ($p < 0.005$). The change in slope directly suggests a change in either the covariance structure of the two symptom domains, a change in the variance of the negative symptoms, or both. This change in slope we present in Fig. 2.

4. Discussion

Simplified PANSS factor models fit better across samples than models which retained all 30 symptoms, when the five-factor structure was imposed on the PANSS. Using a simplified PANSS with an oblique rotation instead of a full PANSS led to a more stable factor structure, yet still this five-factor structure did not generalize across treatment phase; even the “best” models were technically not good enough according to standard fit metrics. This suggests that the methodological framework used to model the PANSS may not be appropriate, as the problems may be with the model family rather than the PANSS itself.

The strong performance of sparse factor models, which removed large portions of the PANSS, suggests that low-loading items are sources of variability because they are not unique to a single factor; these low-loading items need multiple measurements (i.e. cross-loadings) in order to be reproducible across samples. However, the majority of methods reported to date use an assumption that the underlying dimensions are unrelated to each other and that there is no residual correlation, with the notable exception of van der Gaag et al. (2006). This assumption stipulates, for example, that Preoccupation is either a Positive item or Disorganized item, but not both. This is problematic because the loadings of items are inherently continuous, and a binary winner may be chosen based on a statistical tie.

The factor structure differed depending on whether a patient had received a treatment. This problematically implies that the PANSS may capture different constructs across treatment phases, analogous to using a different scale to measure weight at different timepoints. The factor structure changing with treatment phase is supported by the well-established finding in drug trials that positive symptoms tend to be more responsive to treatment. The change in factor structure then does not necessarily indicate a flaw in the PANSS scale; it may not be possible to create an instrument resilient to treatment stage which can also capture a defining feature of schizophrenia: the positive symptoms. However, this does suggest that negative symptoms, which here were similar to the Marder model's original definition, may provide a set of “anchor items” which might be more resilient. Analyses may be improved then by holding the change in total scores constant not with respect to the total PANSS score, which includes positive symptoms, but instead with respect to only the “anchor” items.

The failure of the five-factor model to replicate across studies similarly does not suggest that the correspondence between items and factors does not replicate for any factors, but instead could be driven by the variation in a single factor. This is supported by the consistency of the negative factor across studies. In this study, the negative symptoms observed were remarkably similar to the Marder model, which has been replicated in patients with differing

illness phases (Lancon et al., 2000; Emsley et al., 2003; Mohr et al., 2004; Klingberg et al., 2006).

This study has limitations; the baseline values were captured when most patients were clinically unstable or acutely ill. These results might differ for patients who were clinically stable or being successfully treated with medications, a direction for future work. This is supported by the finding that pathological characteristics differ among first-episode, chronic and ambulatory patients with Schizophrenia (Khan et al., 2014), as well as the finding that the factor structure differed across treatment stage.

Although only simplified PANSS models were reproducible across samples, this does not imply that the omitted symptoms did not contain useful information about the underlying illness, or that a reduced version of the PANSS still measures what it is intends to measure. Rather, this suggests that factorial models of the PANSS need to venture outside the traditional orthogonal five-factor domain, incorporating other dimensions onto which all items could cross-load, such as bifactor models. This study nominates low-loading PANSS items as a source of noise, but in the future we will use item response theory (IRT) (Khan et al., 2011) and bifactor models (Reininghaus et al., 2013) to evaluate the reliability and validity of the PANSS items, and to measure whether the weak items are themselves sample-dependent. These models will help elucidate the dimensions measured by the PANSS, to improve the accuracy and reliability of measuring and detecting, changes in the underlying illness severity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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. Written informed consent was obtained from all participants after the study procedure was fully explained. The authors also thank Monique Franc, Anthony Santos, Nicole Bottrel, William Cafferty and other staff in the Neuroscience Biomarkers of Janssen Research & Development for operational support including anonymization of clinical data.

Supported in part by sponsored research to UCLA by Janssen Research & Development, LLC and NIH – UL1DE019580 and PL1MH083271. Ariana E Anderson, Ph.D., holds a Career Award at the Scientific Interface from BWF.

References

Aggarwal NK, Tao H, Xu K, Stefanovics E, Zhening L, Rosenheck RA. Comparing the PANSS in Chinese and American inpatients: cross-cultural psychiatric analyses of instrument translation and implementation. Schizophrenia Research. 2011; 132:146–152. [PubMed: 21885258]

Anderson A, Chung H, Li Q, Wilcox M, Salvadore G, Wang D, Riese SP, Zhou J, Bilder RM. The Positive and Negative Syndrome Scale Revisited: seeking bifactor dimensions that span schizoaffective. Bipolar and Schizophrenia Diagnoses. 2015 under review.

Bentler PM. Comparative fit indexes in structural models. Psychological Bulletin. 1990; 107:238. [PubMed: 2320703]

Browne, MW.; Cudeck, R.; Bollen, KA.; Long, JS. Alternative ways of assessing model fit. Vol. 154. Sage Focus Editions; 1993. p. 136

Chen F, Curran PJ, Bollen KA, Kirby J, Paxton P. An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. Sociological Methods & Research. 2008; 36:462–494. [PubMed: 19756246]

Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Medicine. 2013; 11:126. [PubMed: 23672542]

Efron, B.; Efron, B. The Jackknife, the Bootstrap and Other Resampling Plans. SIAM; 1982.

Emsley R, Rabinowitz J, Torreman M. The factor structure for the Positive and Negative Syndrome Scale (PANSS) in recent-onset psychosis. Schizophrenia Research. 2003; 61:47–57. [PubMed: 12648735]

Hastie, T.; Tibshirani, R.; Friedman, J.; Hastie, T.; Friedman, J.; Tibshirani, R. The Elements of Statistical Learning. Springer; New York: 2009.

Hayashi N, Igarashi Y, Yamashina M, Suda K. Is there a gender difference in a factorial structure of the Positive and Negative Syndrome Scale? Psychopathology. 2002; 35:28–35. [PubMed: 12006746]

Kass RE, Raftery AE. Bayes factors. Journal of the American Statistical Association. 1995; 90:773–795.

Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. Schizophrenia Bulletin. 1987; 13:261. [PubMed: 3616518]

Kay SR, Sevy S. Pyramidical model of schizophrenia. Schizophrenia Bulletin. 1990; 16:537. [PubMed: 2287938]

Kelley ME, White L, Compton MT, Harvey PD. Subscale structure for the Positive and Negative Syndrome Scale (PANSS): a proposed solution focused on clinical validity. Psychiatry Research. 2013; 205:137–142. [PubMed: 22974521]

Khan A, Lewis C, Lindenmayer JP. Use of Non-parametric Item Response Theory to develop a shortened version of the Positive and Negative Syndrome Scale (PANSS). BMC Psychiatry. 2011; 11:178. [PubMed: 22087503]

Khan A, Lindenmayer JP, Opler M, Kelley ME, White L, Compton M, Gao Z, Harvey PD. The evolution of illness phases in schizophrenia: a non-parametric item response analysis of the Positive and Negative Syndrome Scale. Schizophrenia Research: Cognition. 2014; 1:53–89.

Klingberg PDS, Wittorf A, Wiedemann G. Disorganization and cognitive impairment in schizophrenia: independent symptom dimensions? European Archives of Psychiatry and Clinical Neuroscience. 2006; 256:532–540. [PubMed: 17164977]

Lancon C, Auquier P, Nayt G, Reine G. Stability of the five-factor structure of the Positive and Negative Syndrome Scale (PANSS). Schizophrenia Research. 2000; 42:231–239. [PubMed: 10785581]

Lehoux C, Gobeil MH, Lefèbvre AA, Maziade M, Roy MA. The five-factor structure of the PANSS: a critical review of its consistency across studies. Clinical Schizophrenia & Related Psychoses. 2009; 3:103–110.

Marder SR, Davis JM, Chouinard G. 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. 1997; 58:538–546. [PubMed: 9448657]

Meredith W. Measurement invariance, factor analysis and factorial invariance. Psychometrika. 1993; 58:525–543.

Meyer, P. Psychology. Purdue University; 2003. The cognitive factor of the PANSS: a confirmatory factor analysis and related cognitive correlates.

Mohr PE, Cheng CM, Claxton K, Conley RR, Feldman JJ, Hargreaves WA, Lehman AF, Lenert LA, Mahmoud R, Marder SR. The heterogeneity of schizophrenia in disease states. Schizophrenia Research. 2004; 71:83–95. [PubMed: 15374576]

Payton ME, Greenstone MH, Schenker N. Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance? Journal of Insect Science. 2003; 3

Pornprasertmanit S, Miller P, Schoemann A, Rosseel Y. SemTools: Useful Tools for Structural Equation Modeling. R Package. 2013 available on CRAN.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2013. URL http://www.R-project.org/

Reininghaus U, Priebe S, Bentall RP. Testing the psychopathology of psychosis: evidence for a general psychosis dimension. Schizophrenia Bulletin. 2013; 39:884–895. [PubMed: 22258881]

Revelle W. Psych: Procedures for Psychological, Psychometric, and Personality Research. R package Version 1019. 2011

Rosseel Y. lavaan: An R package for structural equation modeling. Journal of Statistical Software. 2012; 48:1–36.

Van den Oord EJ, Rujescu D, Robles JR, Giegling I, Birrell C, Bukszar J, Murrelle L, Moller HJ, Middleton L, Muglia P. Factor structure and external validity of the PANSS revisited. Schizophrenia Research. 2006; 82:213–223. [PubMed: 16229988]

van der Gaag M, Cuijpers A, Hoffman T, Remijsen M, Hijman R, de Haan L, van Meijel b, van Harten PN, Valmaggia L, de Hert M, Wiersma D. 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. 2006; 85:273–279. [PubMed: 16730430]

Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for Organizational Research. Organizational Research Methods. 2000; 18:4–70.

Wallwork R, Fortgang R, Hashimoto R, Weinberger D, Dickinson D. Searching for a consensus five-factor model of the Positive and Negative Syndrome Scale for schizophrenia. Schizophrenia Research. 2012; 137:246–250. [PubMed: 22356801]

White L, Harvey PD, Opler L, Lindenmayer JP. 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. 1997; 30:263–274. [PubMed: 9353855]

Yang-Wallentin F, Jöreskog KG, Luo H. Confirmatory factor analysis of ordinal variables with misspecified models. Structural Equation Modeling. 2010; 17:392–423.

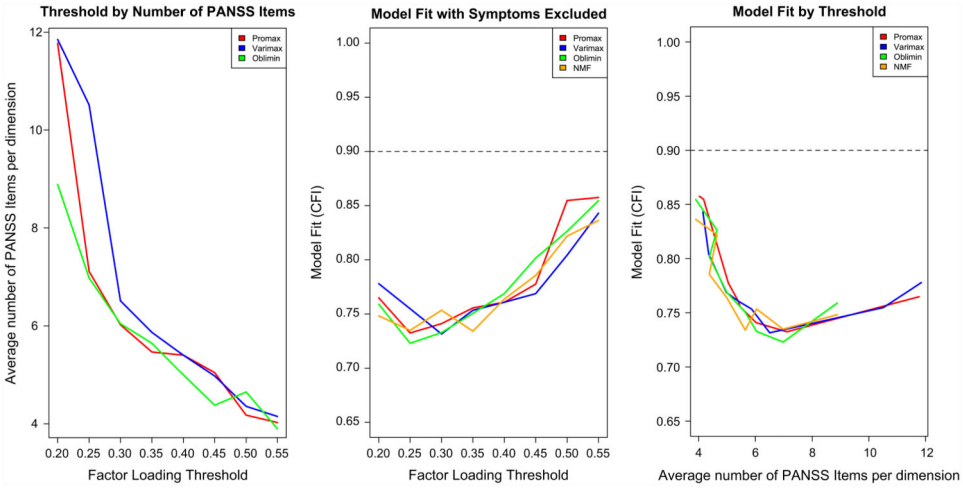

Fig. 1.

(L) As the threshold increased, the number of included PANSS items decreased. (C) Higher thresholds led to better model fits on new samples. (R) Model fit was increased on new samples when fewer items were retained per dimension. Results shown are from baseline observations using a 10-fold cross-validation where a model is trained on 90% of the data, and tested on the remaining 10%, 10 separate times. CFI measures how well the model fit the new data. A higher CFI implies a stronger model fit.

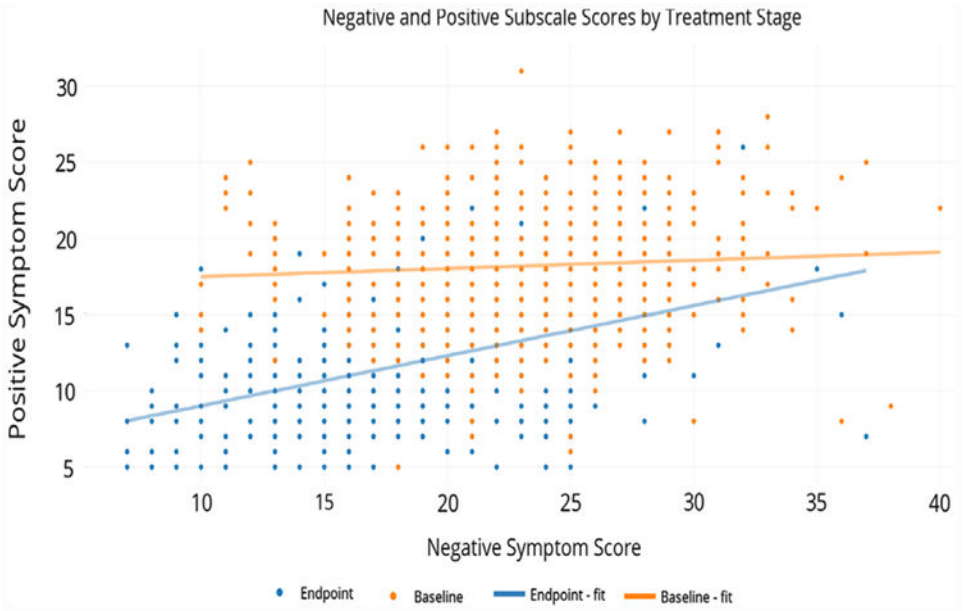

Fig. 2.

After treatment, an increase in negative symptoms was associated with higher scores in positive symptoms (p <0.005). Before treatment, the negative symptoms were less predictive of the positive subscale scores.

Table 1

Study demographics.

ClinicalTrials.gov ID	DX	*N*	Male (%)	Age (S.D.)	Total PANSS (S.D.)
NCT00074477	SZ	328	0.61	39.78 (9.8)	92.1 (11.3)
NCT00334126	SZ	204	0.58	35.92 (10.76)	105.2 (13.9)
NCT00085748	SZ	93	0.26	69.58 (4.56)	92.9 (9.2)
NCT00078039	SZ	555	0.51	37.15 (10.86)	93.6 (10.7)
NCT00077714	SZ	315	0.73	41.81 (10.56)	93.9 (11.8)
NCT00083668	SZ	516	0.66	36.74 (10.54)	92.8 (12.4)
NCT00210717	SZ	576	0.59	40.7 (11.69)	81.1 (13)
NCT00210548	SZ	237	0.67	39.07 (10.36)	90.8 (12.1)
NCT00101634	SZ	372	0.63	40.01 (11.28)	91 (11.9)
NCT00590577	SZ	437	0.66	39.42 (10.7)	87 (11)
NCT00074477	SZ	178	0.66	38.98 (10.47)	87.3 (11.7)
NCT00299715	BP	299	0.54	39.55 (10.86)	64.8 (16)
NCT00309699	BP	350	0.59	40.21 (10.78)	58 (14.3)
NCT00309686	BP	209	0.54	40.79 (11.68)	59.3 (17.6)
NCT00397033	SA	314	0.65	37.22 (10.47)	93.7 (12.8)
NCT00412373	SA	278	0.57	37.47 (9.15)	91.9 (12.6)

Table 2

Averaged 10-fold cross-validation loadings for baseline observations. Loadings >0.55 are highlighted. Retaining symptoms above this threshold led to the most resilient and stable models across samples as measured by the average CFI, yet omitted almost 40% of total PANSS items. The loading matrix is also supplied as a supplementary file.

	Negative	Positive	Anxiety	Excited	Disorganized
Blunted Affect	0.678	−0.063	−0.028	−0.149	0.159
Emotional Withdrawal	0.817	0.056	−0.004	−0.011	−0.002
Poor Rapport	0.611	−0.139	−0.074	0.226	0.234
Passive Apathetic Social Withdrawal	0.813	0.053	−0.009	0.004	−0.063
Lack of Spontaneity and Flow of Conversation	0.601	−0.156	−0.020	0.019	0.216
Motor Retardation	0.520	−0.117	0.140	−0.089	0.032
Active Social Avoidance	0.588	0.234	0.111	0.117	−0.155
Delusions	−0.028	0.837	0.016	0.004	0.067
Hallucinatory Behavior	0.047	0.604	0.071	0.016	−0.096
Grandiosity	−0.250	0.357	−0.110	0.178	0.077
Suspiciousness Persecution	0.136	0.616	0.124	0.186	−0.177
Stereotyped Thinking	0.096	0.113	0.093	−0.043	0.548
Somatic Concern	−0.065	0.048	0.303	−0.039	0.043
Unusual Thought Content	−0.076	0.696	−0.063	−0.016	0.211
Lack of Judgment and Insight	0.094	0.159	−0.221	0.121	0.451
Conceptual Disorganization	0.072	0.252	−0.048	0.020	0.612
Difficulty in Abstract Thinking	0.133	0.094	−0.103	0.010	0.388
Mannerisms and Posturing	0.107	−0.126	0.043	0.052	0.440
Poor Attention	0.061	−0.053	0.103	0.090	0.590
Disturbance of Volition	0.297	−0.040	0.095	0.006	0.436
Preoccupation	0.169	0.349	0.120	−0.049	0.382
Disorientation	0.056	0.038	−0.082	0.144	0.196
Excitement	−0.259	0.066	0.253	0.501	0.213
Hostility	0.026	0.065	0.012	0.849	−0.142
Uncooperativeness	0.126	−0.034	−0.072	0.749	0.091
Poor Impulse Control	−0.140	0.014	0.086	0.655	0.091

Psychiatry Res. Author manuscript; available in PMC 2016 February 28.

	Negative	Positive	Anxiety	Excited	Disorganized
Anxiety	−0.025	0.050	0.798	−0.027	0.006
Guilt Feelings	0.009	0.082	0.446	−0.062	−0.219
Tension	−0.012	−0.045	0.696	0.154	0.196
Depression	0.187	0.019	0.511	−0.070	−0.290

Table 3

Averaged 10-fold cross-validation loadings for post-treatment observations. Loadings >0.55 are highlighted. Retaining symptoms above this threshold led to the most resilient and stable models across samples, yet omitted almost 40% of total PANSS items. The loading matrix is also supplied as a supplementary file.

	Negative	Positive	Anxiety	Excited	Disorganized
Blunted Affect	0.760	−0.020	−0.041	−0.071	0.138
Emotional Withdrawal	0.849	0.080	0.038	0.000	−0.011
Poor Rapport	0.655	−0.086	−0.089	0.307	0.169
Passive Apathetic Social Withdrawal	0.816	0.094	0.042	−0.001	−0.035
Lack of Spontaneity and Flow of Conversation	0.671	−0.085	−0.025	0.063	0.167
Motor Retardation	0.570	−0.066	0.183	−0.076	0.045
Active Social Avoidance	0.579	0.186	0.203	0.068	−0.107
Delusions	0.033	0.910	0.023	0.011	−0.004
Hallucinatory Behavior	0.086	0.657	0.138	0.021	−0.051
Grandiosity	−0.215	0.518	−0.088	0.307	0.098
Suspiciousness Persecution	0.122	0.594	0.175	0.253	−0.195
Stereotyped Thinking	0.172	0.249	0.089	−0.015	0.501
Somatic Concern	−0.001	0.081	0.504	0.000	0.051
Unusual Thought Content	−0.036	0.787	0.019	−0.024	0.172
Lack of Judgment and Insight	0.207	0.299	−0.192	0.186	0.387
Conceptual Disorganization	0.127	0.316	0.009	0.068	0.550
Difficulty in Abstract Thinking	0.237	0.192	−0.051	0.044	0.394
Mannerisms and Posturing	0.200	−0.024	0.028	0.109	0.467
Poor Attention	0.160	−0.015	0.143	0.131	0.597
Disturbance of Volition	0.320	−0.024	0.097	0.033	0.477
Preoccupation	0.195	0.321	0.216	−0.056	0.365
Disorientation	0.159	0.130	−0.027	0.233	0.243
Excitement	−0.192	0.130	0.297	0.512	0.244
Hostility	0.044	0.079	0.096	0.844	−0.141
Uncooperativeness	0.210	0.030	−0.081	0.715	0.097
Poor Impulse Control	−0.109	0.037	0.168	0.661	0.154

	Negative	Positive	Anxiety	Excited	Disorganized
Anxiety	−0.015	0.056	0.822	0.041	0.036
Guilt Feelings	0.023	0.097	0.562	0.011	−0.109
Tension	−0.028	−0.010	0.694	0.176	0.236
Depression	0.218	0.036	0.659	−0.022	−0.186

Table 4

We tested whether PANSS factor models fit differently across treatment stages for different sparsity levels. We created two restricted PANSS models on the pre-treatment and post-treatment data separately, and tested these models on both the pre-treatment and post-treatment stages separately. We compared these fits to the full-factor models: the 30-item Marder model and the 30-item Proposed model which was created using an EFA on the pre-treatment data. All models fit better to the post-treatment data than the pre-treatment data, and sparse factor models fit better than full-factor models.

PANSS Items Retained	PANSS Model	Testing Data	RMSEA	90% CI RMSEA	SRMR	CFI	BIC[a]
18	Pre-treatment	Pre-treatment	0.081	(0.079, 0.084)	0.067	0.761	164,878.30
18	Pre-treatment	Post-treatment	0.082	(0.079, 0.084)	0.055	0.825	145,287.61
20	Post-treatment	Pre-treatment	0.083	(0.081, 0.086)	0.055	0.712	184,725.29
20	Post-treatment	Post-treatment	0.079	(0.077, 0.081)	0.059	0.811	164,449.44
30	Marder	Pre-treatment	0.082	(0.080, 0.083)	0.087	0.581	282,657.41
30	Marder	Post-treatment	0.079	(0.077, 0.080)	0.069	0.713	251,986.79
30	Proposed	Pre-treatment	0.07	(0.069, 0.072)	0.073	0.689	280,676.21
30	Proposed	Post-treatment	0.065	(0.064, 0.067)	0.057	0.802	249,040.99

[a] All fit statistics were obtained using DWLS robust estimation, with the exception of the BIC which requires a likelihood value for computation.

Table 5

The squared multiple correlation of PANSS items increased after treatment, for every item. This suggests that the treatment introduced more regularity in the symptom covariance structure.

Item	Pre-treatment	Post-treatment
Emotional Withdrawal	0.626	0.746
Passive Apathetic Social Withdrawal	0.613	0.717
Delusions	0.613	0.788
Hostility	0.590	0.715
Poor Rapport	0.568	0.677
Blunted Affect	0.547	0.628
Anxiety	0.527	0.698
Uncooperativeness	0.524	0.658
Tension	0.514	0.695
Lack of Spontaneity and Flow of Conversation	0.511	0.596
Suspiciousness Persecution	0.508	0.714
Unusual Thought Content	0.507	0.676
Excitement	0.494	0.695
Poor Impulse Control	0.484	0.628
Conceptual Disorganization	0.452	0.650
Active Social Avoidance	0.449	0.597
Depression	0.421	0.537
Motor Retardation	0.414	0.479
Hallucinatory Behavior	0.391	0.601
Preoccupation	0.377	0.566
Poor Attention	0.365	0.571
Disturbance of Volition	0.349	0.498
Lack of Judgment and Insight	0.348	0.555
Guilt Feelings	0.343	0.424
Stereotyped Thinking	0.342	0.557
Difficulty in Abstract Thinking	0.292	0.477
Grandiosity	0.286	0.480
Mannerisms and Posturing	0.239	0.401
Disorientation	0.233	0.400
Somatic Concern	0.116	0.335

Table 6

CFI measurements of CFA for 30-item PANSS model, by subgroup. Patients were split using the median positive symptom change. Patients who exhibited stronger changes in their positive symptoms had greater variability in the factor structure fit across treatment stage than patients who exhibited weaker changes in positive symptoms.

	Low Positive Subscale	High Positive Subscale
Baseline	0.71	0.692
Post-treatment	0.753	0.822

Table 7

After treatment, the relationship of the positive and negative symptoms change (p <0.005), as seen by the significant interaction effect between negative symptoms and treatment stage.

Covariate	Estimate	Std. error	*t* value	
(Intercept)	22.20506	0.40924	54.26	***
Negative Symptoms	0.20997	0.01751	11.99	***
Treatment Stage	−12.78332	0.44426	−28.77	***
Negative Symptoms: Treatment Stage	0.40528	0.02047	19.8	***