



Published in final edited form as:

*Psychiatry Res.* 2013 January 30; 205(1-2): 137–142. doi:10.1016/j.psychres.2012.08.019.

## Subscale structure for the Positive and Negative Syndrome Scale (PANSS): A proposed solution focused on clinical validity

Mary E. Kelley<sup>a,\*</sup>, Leonard White<sup>b</sup>, Michael T. Compton<sup>c</sup>, and Philip D. Harvey<sup>d</sup>

<sup>a</sup>Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, USA

<sup>b</sup>Pilgrim Psychiatric Center, New York State Office of Mental Health, Brentwood, NY, USA

<sup>c</sup>Department of Psychiatry and Behavioral Sciences, The George Washington University, Washington, DC, USA

<sup>d</sup>Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami FL, USA

### Abstract

Although the items of the positive and negative syndrome scale (PANSS) are ordinal, continuous data methods are consistently used to analyze them. The current study addresses this issue by applying a categorical method and critically examining the ideas of item inclusion and goodness of fit. Data from 1527 subjects was used to test a proposed solution to the factor structure of the PANSS using a categorical factor analytic method. The model was made more generalizable by setting a minimum level of association between the item and the factor, and the results were then compared to existing solutions. The model was also tested for consistency in a first-episode sample. Use of categorical methods indicated similar results to previous analyses; however, it is demonstrated that the strength of the estimates can be unstable when items are shared across factors. The current study demonstrates that solutions can change substantially when a model is over-fitted, and therefore use of measures of fit as the criterion for an acceptable model can mask important relationships and decrease clinical validity.

### Keywords

confirmatory factor analysis (CFA); Positive and Negative Syndrome Scale (PANSS); psychotic symptoms; latent variables; schizophrenia

## 1. Introduction

The Positive and Negative Syndrome Scale (PANSS) is the most widely used measure for the assessment of the symptoms of schizophrenia. Thus, the analysis of PANSS structure is paramount to research on the disease, as item-level analysis is prohibitive and use of total scores can obscure neuropharmacological targets (Kirkpatrick and Fischer, 2006). Using both exploratory and confirmatory methods, the literature shows the most replicated solution

© 2012 Elsevier Ireland Ltd. All rights reserved.

\*corresponding author address: Department of Biostatistics and Bioinformatics, Emory University - Rollins School of Public Health, 1518 Clifton Rd NE, Atlanta, GA 30329, Tel: 404-712-0804, Fax: 404-727-1370, mekelle@emory.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

involves five factors (White et al., 1997, van der Gaag et al., 2006), although the content of the factors varies slightly among studies. In fact, while a large number of samples have been tested using various methods (see van der Gaag et al., 2006), the striking fact about the results as a whole is that the solutions are much more alike than different (Kay and Sevy, 1990, Bell et al., 1994, Lindenmayer et al., 1995, Marder et al., 1997, White et al., 1997, van der Gaag et al., 2006, Citrome et al., 2011, Reininghaus et al., 2012).

The literature also shows that the majority of analyses used to determine underlying scale structure have relied on continuous (normality-based) methods even though the items of the PANSS are ordinal categories. These methods are not without merit because factor analysis has been shown to be relatively robust to violations of normality assumptions such as skewness (Fuller and Hemmerle, 1966). However, it is often overlooked that normality also implies *continuity* (i.e., very few or no ties in the data). The inherently large number of ties within the item-level data results in reduced variance (Blalock, 1976) that subsequently affects the estimates of covariance and correlation, the basis of all factor analytic methods. Therefore, the categorical nature of the data can skew the results of the analysis and lead to erroneous conclusions (Olsson, 1979), which may not have been properly addressed in prior assessments due to the limited availability of *fully categorical data methods* for comparison. This may also be the reason why some previous solutions resulted in items being removed (Kay and Sevy, 1990, White et al., 1997), or why complicated error structures were adapted in order to find numerically acceptable solutions to the empirical data (van der Gaag et al., 2006). One paper on the PANSS made note of the normality issues and proposed using principal components analysis (PCA) (Levine and Rabinowitz, 2007), which does not assume a particular distribution. However, it is unclear to what extent PCA would be affected by the reduced variance associated with ordinal item-level data, given that the covariance matrix used for PCA is calculated using normality-based formulas.

Thus the current study uses a novel categorical analytic technique (Rabe-Hesketh et al., 2004) to fit a proposed factor structure for the PANSS (van der Gaag et al., 2006) using confirmatory factor analysis (CFA). The choice of CFA in this setting is due to the overwhelming commonality among the other previous solutions, and the goal of the current paper is to examine the consistency among solutions rather than generate a new solution. The benefit of the use of categorical methods is that by definition they calculate variance differently by modeling the probabilities of response, and thus do not suffer from the same issues as normality based methods described above. The fit of this large empirical dataset is then used to revisit the issues of item inclusion and subscale structure. We then test this solution on a new, independent dataset from a first-episode psychosis sample to see if the solution could also be used for less chronic patients. The benefits of this strategy are that such a sample allows for an examination of latent structure of symptoms relatively unobscured by chronicity and treatment effects. In contrast, patients in the early stages of illness may have different symptom profiles (either related to severity or the relationship among symptoms), which can also affect the attempts to replicate results using CFA. Even so, several studies using CFA have demonstrated that the structure of PANSS-rated symptoms in first-episode patients may be comparable to that reported in more chronic samples (Drake et al., 2003, Reininghaus et al., 2012).

Finally, we examine the concept of “goodness of fit” as the ultimate goal of CFA; mainly because the specificity of a particular model to the dataset from which it is derived is especially a concern when considerable adjustments to the model are made to improve fit. Such adjustments may degrade the clinical validity of models by either 1) making the solution too sample-specific (i.e., not generalizable) and/or 2) forcing the exclusion of core symptoms of the illness due to lack of variance in the items. We propose an alternative -- namely the standardized “size” of the loading (i.e., the magnitude of the correlation between

the item and the factor) -- as essential evidence for the adequacy of a particular model to the data. The discussion of the size of the loading is often considered in the building of these structural models using exploratory factor analysis (EFA) (Cudeck and Odell, 1994, Skrondal and Rabe-Hesketh, 2004) but is for the most part not discussed when using confirmatory methods. However, the assessment of the strength of the relationships between the item and the factor demonstrates a type of clinical relevance that would likely be crucial to the generalizability of the result to other data sets.

Most importantly, the assumption of the authors is that because every solution is to some extent sample-specific, it might be preferable to adopt a “consensus” model, similar to the current solutions in neuropsychology, rather than to fit new and different models to a multitude of data sets. If the ultimate goal of the determination of factor structure is to define valid and reproducible underlying constructs within a scale, we suggest that with the recent emphasis (and perhaps over-emphasis) on overall model fit, this goal is often overshadowed. Thus, the purpose of this paper is to examine the evidence for consistency across studies, and discuss possible methodological reasons for some previous inconsistencies.

## 2. Methods

### 2.1. Factor analytic methods for ordinal data

The earliest proposed solution for the confirmatory factor analysis of ordinal data was a computational compromise proposed by both Muthén (Muthén, 1984) and later in a slightly different form by Jöreskog (Jöreskog, 1990, Jöreskog, 1994). This underlying variable (UV) approach assumes that each ordinal item is measuring an underlying, unobserved variable that is normally distributed. It is not “fully categorical”, however, as it relies on the assumption of bivariate normality to calculate polychoric correlations which are then in turn used to fit the CFA model in a similar manner to the traditional linear case.

Previous statistical assessments of these UV methods have tested the sensitivity of the methods to skewness and kurtosis (non-normality) of the observed distributions (Potthast, 1993) as well as the underlying distributions (Flora and Curran, 2004). The results have shown that while the parameter estimates (loadings) are somewhat robust to moderate deviations from normality, use of the UV method leads to consistently inflated test statistics and underestimated standard errors, and this bias increases with smaller sample sizes and larger (at least 10–20) numbers of parameters (Potthast, 1993, Flora and Curran, 2004). The effects of non-normality in shape were particularly pronounced in the instance where the data exhibited high positive kurtosis, or what is referred to in other applications as “zero inflation” (Lachenbruch, 2002, Kelley and Anderson, 2008). There is some indication that robust estimation attenuates these effects (Flora and Curran, 2004, Yang-Wallentin et al., 2010) and can be accomplished through either robust WLS (Mplus) or using the asymptotic covariance matrix (LISREL). These two methods provide nearly identical results when all items are ordinal (Yang-Wallentin et al., 2010). Further developments of these UV methods for exploratory factor analysis (EFA) that were not only bivariate, but multivariate (i.e. full likelihood methods) have been proposed (Lee et al., 1990) but were found to be computationally unfeasible for models with more than a few factors (Jöreskog and Moustaki, 2001). The *fully categorical ordinal data method* used in the current investigation is the traditional ordinal logistic regression model assuming proportional odds and the logistic distribution function (McCullagh, 1980), applied to the latent variable model used for CFA (Rabe-Hesketh et al., 2004), implemented by an add-on to Stata ([www.gllamm.org](http://www.gllamm.org)). It is important to note that this full likelihood method differs from item response theory (IRT) (Forero and Maydeu-Olivares, 2009, Reininghaus et al., 2012) logistic models which are confirmatory in nature, but parameterized differently. The current method models the probabilities of each response (k) to each item (m) as a multivariate

vector per subject, rather than the probabilities of response *patterns* or combinations (total possible =  $k^m$ ) across the sample, as in IRT. The attempt to model the patterns has until recently limited the estimation to small numbers of items and very few factors (approximately 2 or 3) (Jöreskog and Moustaki, 2001, Forero and Maydeu-Olivares, 2009), as the computation increases exponentially with the number of items and factors. This limitation, along with the fact that the Stata model is the only fully categorical model analogous in structure to the CFA for normally distributed variables, influenced our choice of the Statabased model over the Mplus IRT competitor.

## 2.2. Data description

**2.2.1 Chronic schizophrenia sample**—The majority of the clinical data (71.6%) for this demonstration was obtained from previous analyses of the PANSS done on a collection of datasets (Kay and Sevy, 1990, Bell et al., 1994, Caton et al., 1994, Caton et al., 1995, Davidson et al., 1995), which resulted in the original 5 factor (Pentagonal) model (White et al., 1997). Briefly, we are using data from four of the five sites from the prior paper, excluding the acute inpatient dataset involving 139 patients, which was not available. In addition, we have new data from two studies of well-characterized ambulatory outpatients with schizophrenia (Bowie et al., 2008, Harvey et al., 2011). Demographics from the samples are listed in Table 1; the sample consisted of 1527 unique subjects with schizophrenia or schizoaffective disorder evaluated upon entry to the study for which they were consented.

**2.2.2 First-episode psychosis sample**—Clinical data for the testing of the model in a first-episode sample came from a combined data set from two observational studies of consecutively hospitalized first-episode psychosis patients in several public-sector settings in Atlanta, Georgia. The two studies primarily examined predictors of treatment delay and the duration of untreated psychosis (Compton et al., 2008, Compton et al., 2009a, Compton et al., 2011) and the impact of premorbid cannabis use on age at onset of psychosis (Compton et al., 2009b), the latter study being ongoing.

## 2.3 Analysis plan

We fit the full model with correlated factors (van der Gaag et al., 2006); however, we did not allow for correlation among the error terms, as these are generally only adopted to improve fit. All factor loadings were standardized to facilitate making recommendations for subscale structure (Bartholomew et al., 2008). For the purposes of “clinical relevance” we chose a somewhat standard (but still arbitrary) correlation of at least 0.30 (Cudeck and Odell, 1994, Skrondal and Rabe-Hesketh, 2004) between the item and the factor, or equivalently, approximately 10% of the variance in the factor being attributable to the item, to indicate reasonable evidence that an item should belong to a particular subscale. Note that this is a minimum threshold, chosen to include some of the less-well used and less variable items of the scale, which often do not reach higher levels of correlation in these solutions (Citrome et al., 2011). This criterion was not used in place of statistical significance but in addition to it; i.e., all loadings identified were also statistically significant (Cudeck and Odell, 1994); in fact with most sample sizes large enough to fit a 30-item model, the two are essentially equivalent. In contrast, the significance of the estimate gives us very little information by which to verify whether an item belongs to a factor, as loadings as small as 0.061 can be “significant.” If the loading is that small, we would argue that there is no empirical evidence that that item should be in that factor, despite what statistical significance of the estimate or goodness of fit measures may indicate.

### 3. Results

#### 3.1. Model fit

The model results are given in Table 2. For comparison, we also fit solutions using traditional linear methods, as well as the UV method (coefficients not shown) and calculated two of the traditional fit indices (comparative fit index, CFI; root mean squared error of approximation, RMSEA). Initial observation shows that overall the models are mostly in agreement; however, the standardized loadings tend to be lower for the ordinal logistic fit. Both of the normality-based model solutions had more than adequate fit (CFI=0.935, 0.957, respectively) and relatively low RMSEA (0.08 and 0.07, respectively). For simplicity, we marked the item loadings which were *NOT* significant in magnitude relative to the standard error in Table 2, as nearly all loadings, even those with small magnitudes, were statistically significant.

#### 3.2. The move toward a generalizable subscale structure

Although the overall fit indices would be judged as mostly adequate, it appears that there are a number of items that we would recommend removing from certain factors due to the lack of evidence of a clinically relevant association [Table 2]. We then fit a “reduced” model removing those items in order to better assess the correct loadings on the other items, as the validity of a model can always be negatively affected by the fitting of spurious predictors. More importantly, over-fitting is especially a concern when the variance is reduced as is the case with ordinal data. There were 3 items (avolition (G13), difficulty with abstract thinking (N5), and somatic concern (G1)) for which both loadings are less than 0.3 in the full model; in those cases we decided to choose the higher of the two loadings and allow the item to remain on that factor, to see if the removal of spurious items altered the solution. Only one item, tension (G4), was shown to be “shared” in the sense that loadings on two factors were both > 0.30, so we fit that item on both factors in the reduced model.

The results of the reduced model are also listed in Table 2. Of note, the removal of items with little evidence of association altered the loading of the tension item (G4) on the excited factor to below our decided threshold level of 0.30 (0.098). Similarly, loadings for a number of the items that were below the threshold in the initial fit of the full model were now well above the 0.30 threshold level of clinical relevance. This indicates that the sharing of items across many factors (potential over-fitting) can mask meaningful associations, and result in unstable estimates. Thus, without any adjustments for “fit,” all 30 items were shown to be associated with at least one factor, both statistically and clinically (i.e., of sufficient magnitude to be considered meaningful). In contrast to previous solutions, however, only one item (conceptual disorganization) fit more than one factor, that item being shared across two factors. Thus, although a good portion of the structure of the van der Gaag solution (van der Gaag et al., 2006) was replicated in the current sample, the use of the threshold for inclusion resulted in a model with fourteen fewer parameters, a substantial reduction.

#### 3.3. Testing of the model in first-episode patients

In order to test whether the solution for chronic patients could also be used for first-episode subjects, we fit the reduced model derived on the chronic schizophrenia patients to our first-episode sample. Because the concept of goodness of fit does not apply to these models, and has been shown to be unrelated to the clinical evidence, we again focus on whether or not the model maintains reasonably substantial correlations between the items and the factors. If we again use the threshold of 0.30, we see that the majority of the items meet that criterion, indicating the model is well replicated in this sample. The exceptions are conceptual disorganization (P2,  $r=0.002$ ), and two items of the disorganization factor; i.e., mannerisms and posturing (G5,  $r=0.126$ ) and disorientation (G10,  $r=0.212$ ). However, we would argue

that only the loading of conceptual disorganization (P2) shows clear evidence that the item may not belong in the negative factor. Thus, given that it is the only remaining item shared among factors in our ordinal analysis, we chose to drop it from the negative factor in the proposed model; this was also in agreement with many of the previously derived models. Although the items in the disorganization factor have low correlations, they do not provide strong evidence of lack of fit of the proposed model for this sample of first-episode patients, given the reduced sample size, and that they are in the proposed direction. Further, if we agree that the ultimate goal is to create subscales that can be used to enhance comparability of research results across future studies, a small number of items will hardly affect the subscale totals, and we must expect some necessary variation among samples due to random sources and sample characteristics.

### 3.4. Proposal of a consensus subscale structure

We propose a subscale structure to be used for future research in Table 2, consisting of those items that remained clinically significant for our data set, and also fit well with the first-episode data. Interestingly, with the exception of only a few items (one on the disorganized factor, and one on the emotion factor) these items are identical to those items indicated to have been found in 9/10 of the cross validations in the van der Gaag study (van der Gaag et al., 2006). Thus, it would appear that our criterion of a standardized loading of  $> 0.30$  (for clinical significance) may have had a similar effect as the cross-validation in reducing spurious associations due to outliers in the data set, and is much easier to implement. Furthermore, the reduced model revealed here is also remarkably similar to a number of other models (Table 3), including both the original 4-factor pyramidal model (Kay and Sevy, 1990), the well-replicated pentagonal model (White et al., 1997), a model derived from the North American trials of risperidone (Marder et al., 1997), and a more recent large-sample analysis of seven trials of iloperidone (Citrome et al., 2011). Most significantly, it is quite similar to another more recent model derived using slightly different categorical factor analytic methods (Reininghaus et al., 2012). That study illustrated that the proposed structure remains even in the presence of a “general” factor, thus it would appear the model has considerable face validity.

## 4. Discussion

It is important to note that we are not the first group to recognize the similarity across the solutions, and look for ways to combine results (Lehoux et al., 2009, Wallwork et al., 2012). The novelty of the current study is in addressing the categorical nature of the data using new methods and in the discussion of the rules for item inclusion. The fact that the final solution presented here is consistent with many previous studies provides evidence that the normality-based measures are not fatally flawed in regards to identifying the major factors within the PANSS scale, and that the proposed solution is reasonable to use for studies of chronic patients.

The use of the model in first-episode patients seems to be mostly supported by the data, but further replication across additional samples would be needed before general recommendations could be made. More specifically, we acknowledge that the differences reflected in the chronic and first-episode sample could be due to the nature of the early stages of schizophrenia, given that symptoms such as disorientation and conceptual disorganization are commonly associated with later phases of the disease, and could also be less prominent in later cohorts treated with newer antipsychotics. In addition, there may be differences in clinical and comorbidity profiles across these samples which account for these differences; unfortunately we do not have the data to formally make comparisons.

In contrast, the current approach did uncover something that was not previously addressed, which is the effect of over-fitting, which is the direct result of the pursuit of a high goodness of fit index (GFI). Three items were not meaningfully associated in the full model but became so in the reduced model, after removing non-correlated items. Most notably, difficulty with abstract thinking (N5) changed substantially, indicating that the effects of over-fitting can be severe.

Despite the fact that the categorical solution is less familiar to some researchers, it may be important to adopt as many of the items on the PANSS are not only skewed, but have very limited variance altogether. A case in point is the item of conceptual disorganization (P2) – both the pyramidal and pentagonal models dropped the item, and the van der Gaag solution had it as a negative loading in the negative factor (van der Gaag et al., 2006), which would appear to be contradictory to the proposed model. This is most likely due to the fact that the item is not continuous, e.g. in the current data 32% of the sample does not exhibit this symptom. In fact, the majority of items on the PANSS exhibit strikingly similar patterns, thus the use of the categorical analysis would be preferable for this scale due to the fact that it models the probabilities of each response category, which allows them to vary substantially, i.e. it does not require normality in the shape of the distribution. Unfortunately, the categorical solution illustrated here requires substantial time computationally; it appears, however, that the use of the IRT solution (i.e. Mplus) provides similar results (Reininghaus et al., 2012) without such a time constraint, which is encouraging.

Finally, we recommend the use of a threshold for the item-to-factor correlation in choosing items for subscales, rather than relying on statistical significance and goodness of fit indices, regardless of the method of estimation used to fit a solution. Our use of the magnitude of the standardized loadings could be considered analogous to the issue of clinical significance versus statistical significance argued in the discussion of clinical trial results (Kraemer, 2006, Kraemer and Kupfer, 2006). However, this is rarely discussed in the context of factor analysis. Indices of fit can certainly be useful; however, we would propose that it is the dependence on these measures while excluding other important criteria that is the concern raised by the current investigation. While benchmarks for GFI (e.g., 0.90) and RMSEA are for the most part consistently used, they are also arbitrary and do not reflect clinical relevance or reproducibility of a model; the same could be said for traditional tests of model significance such as likelihood ratio tests which apply to the categorical models. Notably, for either the linear or UV fit of the current model, the results would have been considered a successful replication and no adjustments would have been made. If the goal is to use the results of the CFA to provide recommendations for the creation of subscales for future use, it would appear that not only do the fit indices not provide sufficient information in this regard, but they could possibly be misleading as we can always increase the fit by altering the model. We suggest that for the purposes of validation of subscale structure, the magnitude of the association should be used as an important indicator of clinical relevance along with standard estimates of statistical significance.

## Acknowledgments

Dr. Kelley would like to thank the principal investigators who provided data for the current analysis. The study was funded in part by NIMH Grants 78775 and 63116 to Dr. Harvey, and grant R01 MH081011 to Dr. Compton.

## References

Bartholomew, DJ.; Steele, F.; Moustaki, I.; Galbraith, JI. Analysis of Multivariate Social Science Data. 2nd ed.. Boca Raton: CRC Press; 2008.

- Bell MD, Lysaker PH, Beam-Goulet JL, Milstein RM, Lindenmayer JP. Five-Component Model of Schizophrenia: Assessing the Factorial Invariance of the Positive and Negative Syndrome Scale. *Psychiatry Research*. 1994; 52(3):295–303. [PubMed: 7991723]
- Blalock HM Jr. Can We Find a Genuine Ordinal Slope Analogue? *Sociological Methodology*. 1976:7195–7229.
- Bowie CR, Leung WW, Reichenberg A, McClure MM, Patterson TL, Heaton RK, Harvey PD. Predicting Schizophrenia Patients' Real-World Behavior with Specific Neuropsychological and Functional Capacity Measures. *Biological Psychiatry*. 2008; 63(5):505–511. [PubMed: 17662256]
- Caton CL, Shrout PE, Dominguez B, Eagle PF, Opler LA, Cournos F. Risk Factors for Homelessness among Women with Schizophrenia. *American Journal of Public Health*. 1995; 85(8 Pt 1):1153–1156. [PubMed: 7625518]
- Caton CL, Shrout PE, Eagle PF, Opler LA, Felix A, Dominguez B. Risk Factors for Homelessness among Schizophrenic Men: A Case-Control Study. *American Journal of Public Health*. 1994; 84(2):265–270. [PubMed: 8296951]
- Citrome L, Meng X, Hochfeld M. Efficacy of Iloperidone in Schizophrenia: A Panss Five-Factor Analysis. *Schizophr Res*. 2011; 131(1–3):75–81. [PubMed: 21700430]
- Compton MT, Chien VH, Leiner AS, Goulding SM, Weiss PS. Mode of Onset of Psychosis and Family Involvement in Help-Seeking as Determinants of Duration of Untreated Psychosis. *Soc Psychiatry Psychiatr Epidemiol*. 2008; 43(12):975–982. [PubMed: 18604616]
- Compton MT, Gordon TL, Goulding SM, Esterberg ML, Carter T, Leiner AS, Weiss PS, Druss BG, Walker EF, Kaslow NJ. Patient-Level Predictors and Clinical Correlates of Duration of Untreated Psychosis among Hospitalized First-Episode Patients. *J Clin Psychiatry*. 2011; 72(2):225–232. [PubMed: 21382306]
- Compton MT, Goulding SM, Gordon TL, Weiss PS, Kaslow NJ. Family-Level Predictors and Correlates of the Duration of Untreated Psychosis in African American First-Episode Patients. *Schizophr Res*. 2009a; 115(2–3):338–345. [PubMed: 19833482]
- Compton MT, Kelley ME, Ramsay CE, Pringle M, Goulding SM, Esterberg ML, Stewart T, Walker EF. Association of Pre-Onset Cannabis, Alcohol, and Tobacco Use with Age at Onset of Prodrome and Age at Onset of Psychosis in First-Episode Patients. *American Journal of Psychiatry*. 2009b; 166(11):1251–1257. [PubMed: 19797432]
- Cudeck R, Odell LL. Applications of Standard Error-Estimates in Unrestricted Factor-Analysis - Significance Tests for Factor Loadings and Correlations. *Psychological Bulletin*. 1994; 115(3):475–487. [PubMed: 8016288]
- Davidson M, Harvey PD, Powchik P, Parrella M, White L, Knobler HY, Losonczy MF, Keefe RS, Katz S, Frecska E. Severity of Symptoms in Chronically Institutionalized Geriatric Schizophrenic Patients. *American Journal of Psychiatry*. 1995; 152(2):197–207. [PubMed: 7840352]
- Drake RJ, Dunn G, Tarrier N, Haddock G, Haley C, Lewis S. The Evolution of Symptoms in the Early Course of Non-Affective Psychosis. *Schizophr Res*. 2003; 63(1–2):171–179. [PubMed: 12892871]
- Flora DB, Curran PJ. An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis with Ordinal Data. *Psychological Methods*. 2004; 9(4):466–491. [PubMed: 15598100]
- Forero CG, Maydeu-Olivares A. Estimation of Irt Graded Response Models: Limited Versus Full Information Methods. *Psychological Methods*. 2009; 14(3):275–299. [PubMed: 19719362]
- Fuller EL, Hemmerle WJ. Robustness of Maximum-Likelihood Estimation Procedure in Factor Analysis. *Psychometrika*. 1966; 31(2):255–255. [PubMed: 5222212]
- Harvey PD, Raykov T, Twamley EW, Vella L, Heaton RK, Patterson TL. Validating the Measurement of Real-World Functional Outcomes: Phase I Results of the Valero Study. *American Journal of Psychiatry*. 2011; 168(11):1195–1201. [PubMed: 21572166]
- Jöreskog KG. New Developments in Lisrel: Analysis of Ordinal Variables Using Polychoric Correlations and Weighted Least Squares. *Quality and Quantity*. 1990:24387–24404.
- Jöreskog KG. On the Estimation of Polychoric Correlations and Their Asymptotic Covariance-Matrix. *Psychometrika*. 1994; 59(3):381–389.
- Jöreskog KG, Moustaki I. Factor Analysis of Ordinal Variables: A Comparison of Three Approaches. *Multivariate Behavioral Research*. 2001; 36(3):347–387.



- Kay SR, Sevy S. Pyramidal Model of Schizophrenia. *Schizophr Bull.* 1990; 16(3):537–545. [PubMed: 2287938]
- Kelley ME, Anderson SJ. Zero Inflation in Ordinal Data: Incorporating Susceptibility to Response through the Use of a Mixture Model. *Statistics in Medicine.* 2008; 27(18):3674–3688. [PubMed: 18351711]
- Kirkpatrick B, Fischer B. Subdomains within the Negative Symptoms of Schizophrenia: Commentary. *Schizophr Bull.* 2006; 32(2):246–249. [PubMed: 16492798]
- Kraemer HC. Correlation Coefficients in Medical Research: From Product Moment Correlation to the Odds Ratio. *Statistical Methods in Medical Research.* 2006; 15(6):525–545. [PubMed: 17260922]
- Kraemer HC, Kupfer DJ. Size of Treatment Effects and Their Importance to Clinical Research and Practice. *Biological Psychiatry.* 2006; 59(11):990–996. [PubMed: 16368078]
- Lachenbruch PA. Analysis of Data with Excess Zeros. *Statistical Methods in Medical Research.* 2002; 11(4):297–302. [PubMed: 12197297]
- Lee SY, Poon WY, Bentler P. Full Maximum Likelihood Analysis of Structural Equation Models with Polytomous Variables. *Statistics and Probability Letters.* 1990:991–997.
- Lehoux C, Gobeil M-H, Lefebvre A-A, Maziade M, Roy M-A. The Five Factor Structure of the Panss: A Critical Review of Its Consistency across Studies. *Clinical Schizophrenia and Related Psychoses.* 2009; 3(2):103–110.
- Levine SZ, Rabinowitz J. Revisiting the 5 Dimensions of the Positive and Negative Syndrome Scale. *Journal of Clinical Psychopharmacology.* 2007; 27(5):431–436. [PubMed: 17873671]
- Lindenmayer JP, Grochowski S, Hyman RB. Five Factor Model of Schizophrenia: Replication across Samples. *Schizophr Res.* 1995; 14(3):229–234. [PubMed: 7766534]
- 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(12):538–546. [PubMed: 9448657]
- McCullagh P. Regression Models for Ordinal Data (with Discussion). *Journal of the Royal Statistical Society B.* 1980; 42(2):109–142.
- Muthén B. A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators. *Psychometrika.* 1984; 49(1):115–132.
- Olsson U. Robustness of Factor-Analysis against Crude Classification of the Observations. *Multivariate Behavioral Research.* 1979; 14(4):485–500.
- Pothast MJ. Confirmatory Factor-Analysis of Ordered Categorical Variables with Large Models. *British Journal of Mathematical & Statistical Psychology.* 1993:46273–46286.
- Rabe-Hesketh S, Skrondal A, Pickles A. Generalized Multilevel Structural Equation Modeling. *Psychometrika.* 2004; 69(2):167–190.
- Reininghaus U, Priebe S, Bental RP. Testing the Psychopathology of Psychosis: Evidence for a General Psychosis Dimension. *Schizophr Bull.* 2012
- Skrondal, A.; Rabe-Hesketh, S. *Generalized Latent Variable Modeling : Multilevel, Longitudinal, and Structural Equation Models.* Boca Raton: Chapman & Hall/CRC; 2004.
- van der Gaag M, Hoffman T, Remijnsen M, Hijman R, de Haan L, van Meijel B, van Harten PN, Valmaggia L, de Hert M, Cuijpers A, Wiersma D. The Five-Factor Model of the Positive and Negative Syndrome Scale Ii: A Ten-Fold Cross-Validation of a Revised Model. *Schizophr Res.* 2006; 85(1–3):280–287. [PubMed: 16730429]
- Wallwork RS, Fortgang R, Hashimoto R, Weinberger DR, Dickinson D. Searching for a Consensus Five-Factor Model of the Positive and Negative Syndrome Scale for Schizophrenia. *Schizophr Res.* 2012
- 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(5):263–274. [PubMed: 9353855]
- Yang-Wallentin F, Joreskog KG, Luo H. Confirmatory Factor Analysis of Ordinal Variables with Misspecified Models. *Structural Equation Modeling.* 2010:17392–17423.

\$watermark-text

\$watermark-text

\$watermark-text

**Table 1**

Clinical description of data used in the investigation

|              | Chronic patient sample |                         |                                  |                       |                    |                     |             | First-episode sample |
|--------------|------------------------|-------------------------|----------------------------------|-----------------------|--------------------|---------------------|-------------|----------------------|
| Series       | Kay & Sevy 1990        | Caton et al., 1994;1995 | Bell et al., 1994                | Davidson et al., 1995 | Bowie et al., 2008 | Harvey et al., 2010 | Total       | Compton              |
| setting      | Inpatient              | Urban community         | Veterans hospital rehabilitation | Geriatric inpatient   | Outpatient         | Outpatient          |             |                      |
| n            | 239                    | 400                     | 150                              | 305                   | 238                | 195                 | 1527        | 200                  |
| Age mean(sd) | 33.1 (10.2)            | 38.8 (10.6)             | 40.2 (8.6)                       | 75.7 (7.0)            | 56.6 (9.7)         | 44.0 (5.2)          | 48.9 (15.1) | 23.6 (4.9)           |
| % male       | 77                     | 50                      | 95                               | 44                    | 73                 | 69                  | 64          | 73                   |

**Table 2**

Standardized item loadings for confirmatory factor analysis (CFA) of the proposed (van der Gaag, 2006) model to the chronic data (n=1527) and first episode data (n=200)<sup>^</sup>

| Factor 1:<br>negative (NEG)     | Item                                   | Ordinal<br>logistic    | Reduced<br>model | First<br>episode |
|---------------------------------|--|------------------------|------------------|------------------|
|                                 | Conceptual disorganization (P2)        | 0.371                  | 0.308            | 0.002            |
|                                 | Blunted affect (N1)                    | 0.736                  | 0.714            | 0.701            |
|                                 | Emotional withdrawal (N2)              | 0.760                  | 0.743            | 0.745            |
|                                 | Poor rapport (N3)                      | 0.608                  | 0.637            | 0.567            |
|                                 | Social apathy (N4)                     | 0.769                  | 0.753            | 0.784            |
|                                 | Lack of spontaneity (N6)               | 0.738                  | 0.715            | 0.584            |
|                                 | Motor retardation (G7)                 | 0.497                  | 0.457            | 0.341            |
|                                 | Uncooperativeness (G8)                 | 0.084 <sup>+</sup>     | ---              | ---              |
|                                 | Avolition (G13)                        | 0.188 <sup>&amp;</sup> | 0.416            | 0.445            |
|                                 | Active social avoidance (G16)          | 0.439                  | 0.612            | 0.712            |
| Factor 2:<br>positive (POS)     |  |                        |                  |                  |
|                                 | Delusions (P1)                         | 0.796                  | 0.781            | 0.886            |
|                                 | Hallucinations (P3)                    | 0.701                  | 0.680            | 0.806            |
|                                 | Grandiosity (P5)                       | 0.574                  | 0.480            | 0.611            |
|                                 | Suspiciousness (P6)                    | 0.630                  | 0.712            | 0.848            |
|                                 | Difficulty w/abstract thinking (N5)    | 0.000 <sup>*,+</sup>   | ---              | ---              |
|                                 | Somatic concern (G1)                   | 0.140 <sup>+</sup>     | ---              | ---              |
|                                 | Unusual thought content (G9)           | 0.732                  | 0.697            | 0.644            |
|                                 | Lack of insight (G12)                  | 0.117 <sup>+</sup>     | ---              | ---              |
|                                 | Active social avoidance (G16)          | 0.060 <sup>+</sup>     | ---              | ---              |
| Factor 3:<br>disorganized (DIS) |  |                        |                  |                  |
|                                 | Conceptual disorganization (P2)        | 0.842                  | 0.822            | 0.674            |
|                                 | Difficulty with abstract thinking (N5) | 0.069 <sup>&amp;</sup> | 0.872            | 0.680            |
|                                 | Stereotyped thinking (N7)              | 0.562                  | 0.529            | 0.553            |
|                                 | Mannerisms and posturing (G5)          | 0.398                  | 0.356            | 0.126            |
|                                 | Unusual thought content (G9)           | 0.005 <sup>*,+</sup>   | ---              | ---              |
|                                 | Disorientation (G10)                   | 0.599                  | 0.563            | 0.212            |
|                                 | Poor attention (G11)                   | 0.665                  | 0.640            | 0.540            |
|                                 | Lack of insight (G12)                  | 0.814                  | 0.840            | 0.869            |
|                                 | Avolition (G13)                        | 0.171 <sup>+</sup>     | ---              | ---              |

| Factor 1:<br>negative (NEG)         | Item                             | Ordinal<br>logistic    | Reduced<br>model   | First<br>episode |
|-------------------------------------|----------------------------------|------------------------|--------------------|------------------|
|                                     | Preoccupation (G15)              | 0.614                  | 0.604              | 0.604            |
| Factor 4: excited<br>(EXC)          | Item                             | Ordinal<br>logistic    | Reduced<br>model   | First<br>episode |
|                                     | Excitement (P4)                  | 0.528                  | 0.493              | 0.446            |
|                                     | Grandiosity (P5)                 | 0.036 <sup>+</sup>     | ---                | ---              |
|                                     | Hostility (P7)                   | 0.531                  | 0.504              | 0.628            |
|                                     | Poor rapport (N3)                | 0.087 <sup>+</sup>     | ---                | ---              |
|                                     | Tension (G4)                     | 0.383                  | 0.098 <sup>+</sup> | ---              |
|                                     | Uncooperativeness (G8)           | 0.325                  | 0.418              | 0.501            |
|                                     | Poor impulse control (G14)       | 0.496                  | 0.468              | 0.628            |
|                                     | Active social avoidance<br>(G16) | 0.035 <sup>+</sup>     | ---                | ---              |
| Factor 5: emotional<br>distress EMO |                                  |                        |                    |                  |
|                                     | Suspiciousness (P6)              | 0.217 <sup>+</sup>     | ---                | ---              |
|                                     | Somatic concern (G1)             | 0.293 <sup>&amp;</sup> | 0.463              | 0.390            |
|                                     | Anxiety (G2)                     | 0.669                  | 0.648              | 0.717            |
|                                     | Guilt feelings (G3)              | 0.459                  | 0.430              | 0.385            |
|                                     | Tension (G4)                     | 0.383                  | 0.360              | 0.372            |
|                                     | Depression (G6)                  | 0.626                  | 0.593              | 0.703            |
|                                     | Preoccupation (G15)              | 0.018 <sup>+</sup>     | ---                | ---              |
|                                     | Active social avoidance<br>(G16) | 0.180 <sup>+</sup>     | ---                | ---              |

<sup>^</sup> all loadings were statistically significant ( $|estimate/se(estimate)| < 1.96$ ) except those marked with \*

<sup>+</sup> suggested items to be removed from the model solution due to little or no association (standardized loading < 0.30)

<sup>&</sup> suggested items to be removed from the model solution due to little or no association (standardized loading < 0.30), but left in to make use of all items

--- not applicable

**Table 3**

Construction of proposed solution and comparison to previous models

|   | # items shared with proposed consensus model                     |  |   |                              |                               |                                   |
|---|--|--|---|------------------------------|-------------------------------|-----------------------------------|
| Proposed model factors and items:       | Van der Gaag et al, 2006 (30 item; 9/10 cross validations model) | Pyramidal Model (25 item; dropped items P2, P6, N5, N7, G10) | Pentagonal model (25 item; dropped items P2, P6, G10, G12, G16) | Marder et al, 1997 (30 item) | Citrome et al, 2011 (30 item) | Reininghaus et al. 2012 (30 item) |
| Sample size                             | 5769   | 240  | 1233  | 512                          | 3580                          | 816                               |
| POS: P1, P3, P5, P6, G9                 | 5/5 (100%)   | 4/5 (80%)  | 4/5 (80%)   | 5/5 (100%)                   | 5/5 (100%)                    | 5/5 (100%)                        |
| NEG: N1, N2, N3, N4, N6, G7, G13, G16   | 8/8 (88%)  | 8/8 (100%)   | 7/8 (88%)   | 7/8 (88%)                    | 7/8 (88%)                     | 8/8 (100%)                        |
| DIS: P2, N5, N7, G5, G10, G11, G12, G15 | 7/8 (63%)  | 3/8 (not in proposed solution – 38%)                         | Aut preocc: 4/8 (50%)   | 7/8(88%)                     | 8/8 (100%)                    | 7/8 (88%)                         |
| EXC: P4, P7, G8, G14                    | 4/4 (100%)   | 4/4 (100%)   | 4/4 (100%)  | 4/4 (100%)                   | 4/4 (100%)                    | 4/4 (100%)                        |
| EMO: G1, G2, G3, G4, G6                 | 4/5 (80%)  | 4/5 (80%)  | Dysphoric: 5/5 (100%)   | 4/5 (80%)                    | 5/5 (100%)                    | 5/5 (100%)                        |