Psychiatry Research 205 (2013) 137–142

Contents lists available at SciVerse ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

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

Mary E. Kelley [a,*], Leonard White [b], Michael T. Compton [c], Philip D. Harvey [d]

[a] Department of Biostatistics and Bioinformatics, Emory University, Rollins School of Public Health, 1518 Clifton Road NE, Atlanta, GA 30329, USA
[b] Pilgrim Psychiatric Center, New York State Office of Mental Health, Brentwood, NY, USA
[c] Department of Psychiatry and Behavioral Sciences, The George Washington University, Washington, DC, USA
[d] Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, FL, USA

ARTICLE INFO

Article history:
Received 29 March 2012
Received in revised form
8 August 2012
Accepted 22 August 2012

Keywords:
Confirmatory factor analysis (CFA)
Positive and Negative Syndrome Scale (PANSS)
Psychotic symptoms
Latent variables
Schizophrenia

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 were 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.

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). Based on both exploratory and confirmatory methods, the literature shows the most replicated solution 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 Jr., 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 explain 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 article 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

* Corresponding author. Tel.: +1 404 712 0804; fax: +1 404 727 1370.
E-mail address: mekelle@emory.edu (M.E. Kelley).

0165-1781/$ - see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.psychres.2012.08.019

(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 article 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 the normality-based methods described above. The fit of this large empirical data set is used to revisit the issues of item inclusion and subscale structure. We then test this solution on a new, independent data set 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 article 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 (1984) and later in a slightly different form by Jöreskog (1990, 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). 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 Stata-based 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., 1994,1995; 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

Table 1
Clinical description of data used in the investigation.

	Chronic patient sample							First-episode sample
Series	Kay and Sevy (1990)	Caton et al. (1994, 1995)	Bell et al. (1994)	Davidson et al. (1995)	Bowie et al. (2008)	Harvey et al. (2011)	Total	Compton
Setting	Inpatient	Urban community	Veterans hospital rehabilitation	Geriatric inpatient	Outpatient	Outpatient		
n	239	400	150	305	238	195	1527	200
Age, mean (S.D.)	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