Psychiatry Research 206 (2013) 75–80

Contents lists available at SciVerse ScienceDirect

Psychiatry Research

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

Using multidimensional modeling to combine self-report symptoms with clinical judgment of schizotypy

Stéphanie M. van den Berg [a,*], Muirne C.S. Paap [a], Eske M. Derks [b,1], Genetic Risk and Outcome of Psychosis (GROUP) investigators[2]

[a] University of Twente, Department of Research Methodology, Measurement, and Data-Analysis, Behavioral Sciences, De zul, P.O. Box 217, 7500 AE Enschede, The Netherlands
[b] University Medical Center Utrecht, Rudolf Magnus Institute of Neuroscience, Department of Psychiatry, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 27 April 2012
Received in revised form
27 August 2012
Accepted 5 September 2012

Keywords:
Item Response Theory (IRT)
Assessment
Measurement invariance
Liability

ABSTRACT

This study investigated psychometric properties of two widely used instruments to measure subclinical levels of psychosis, the Community Assessment of Psychic Experiences (CAPE) and the Structured Interview for Schizotypy-Revised (SIS-R), and aimed to enhance measurements through the use of multidimensional measurement models. Data were collected in 747 siblings of schizophrenia patients and 341 healthy controls. Multidimensional Item-Response Theory, Mokken Scale and ordinal factor analyses were performed. Both instruments showed good psychometric properties and were measurement invariant across siblings and controls. The latent traits measured by the instruments show a correlation of 0.62 in siblings and 0.47 in controls. Multidimensional modeling resulted in smaller standard errors for SIS-R scores. By exploiting correlations among related traits through multidimensional models, scores from one diagnostic instrument can be estimated more reliably by making use of information from instruments that measure related traits.

1. Introduction

Subclinical psychotic experiences are prevalent in the general population (van Nierop et al., 2012; van Os et al., 2009). Even though they rarely transit into a clinical diagnosis of schizophrenia (prevalence 0.5–1%; McGrath et al., 2004), there is evidence for a familial (genetic) continuity between subclinical psychotic experiences and clinical psychotic symptoms (Kendler and Walsh, 1995; Hanssen et al., 2003; van Nierop et al., 2012; Lataster et al., 2009). Subjects diagnosed with a psychotic disorder, such as schizophrenia, may be at the extreme high end of the liability distribution and score above the disease threshold, while subjects who score just below the disease threshold are not diagnosed with a psychotic disorder but may likely develop such a disorder in the future (Bak et al., 2003; Dominguez et al., 2010). The same may be true for other symptoms associated with schizophrenia, which can reveal themselves on the cognitive, interpersonal and emotional level (see e.g. Lenzenweger, 2010). Lenzenweger (2010) refers to the underlying liability for schizophrenia as "schizotypy". It should be noted that some authors use the terms "schizotypy" and "subclinical psychosis" interchangeably. In this article, we use the term "schizotypy" as an overarching construct, including both "positive" and "negative" symptoms. Schizotypical symptoms can be measured in different ways. This study aims to show how information from two widely used screening instruments for schizotypy, one based on a psychiatric interview, the other based on a self-report questionnaire, can be combined using modern statistical techniques, resulting in increased measurement precision.

We focus on the Structured Interview for Schizotypy-Revised (SIS-R) and the Community Assessment of Psychic Experiences (CAPE). These instruments show good test–retest reliability and good inter-rater agreement (Kendler et al., 1989; Vollema and Ormel, 2000;

* Corresponding author. Tel.: +31 53 489 2422x3616.
E-mail addresses: stephanie.vandenberg@utwente.nl, m.c.s.paap@utwente.nl (S.M. van den Berg), e.m.derks@amc.uva.nl (E.M. Derks).
[1] Present address: Academic Medical Centre University of Amsterdam, Department of Psychiatry, PO Box 22660, 1100 DD Amsterdam, The Netherlands.
[2] GROUP investigators are: René S. Kahn, MD, PhD, Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, the Netherlands; Don H. Linszen, MD, PhD, Department of Psychiatry, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands; Jim van Os, MD, PhD, South Limburg Mental Health Research and Teaching Network, EURON, Maastricht University Medical Centre, Maastricht, the Netherlands, and King's College London, King's Health Partners, Department of Psychosis Studies, Institute of Psychiatry, London, England; Durk Wiersma, PhD, Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands; Richard Bruggeman, MD, PhD, Department of Psychiatry, University Medical Center Groningen, University of Groningen; Wiepke Cahn, MD, PhD, Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht; Lieuwe de Haan, MD, PhD, Department of Psychiatry, Academic Medical Centre, University of Amsterdam; Lydia Krabbendam, PhD, South Limburg Mental Health Research and Teaching Network, EURON, Maastricht University Medical Centre; and Inez Myin-Germeys, PhD, South Limburg Mental Health Research and Teaching Network, EURON, Maastricht University Medical Centre.

0165-1781/$ - see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.psychres.2012.09.015

Konings et al., 2006). Konings et al. (2006) showed that the positive and negative dimensions correspond closely to the positive and negative dimensions of the SIS-R. Details on the factorial structure of CAPE and studies on its reliability and validity can be found elsewhere (Stefanis et al., 2002; Brenner et al., 2007; Konings et al., 2006 and citations therein). Previous studies suggest a multifactorial structure for symptoms associated with schizotypy (e.g., Vollema and Hoijtink, 2000; Kendler et al, 1991). However, very few studies involving the CAPE or SIS-R have thus far employed modern test theory (Item Response Theory, IRT), whereas its advantages are being increasingly recognized. Also within the field of psychiatry, the popularity of IRT has been on the rise, both for analyzing the psychometric properties of questionnaires (e.g., Egberink and Meijer, 2011; Paap et al., 2011b), as well as scrutinizing formal diagnoses (Langenbucher et al., 2004; Paap et al., 2011a). IRT provides a conceptual and statistical framework for studying the internal structure of a scale, possible violations of measurement invariance across subpopulations, and measurement precision across trait level (Reise and Waller, 2009). Moreover, it allows the assessment of correlated traits using multidimensional measurement models.

Our main aim is to enhance the estimation of SIS-R scores by using information contained in the correlation between SIS-R and CAPE scores, through the use of multidimensional IRT (MIRT) models. Briefly, MIRT models are IRT models where several latent traits are related to a fairly large number of items, where these latent traits are allowed to be correlated (Reckase, 2009). As psychopathological items are usually endorsed by relatively few healthy individuals, it is difficult if not impossible to distinguish among individuals with medium or low trait levels. This is reflected in the large number of healthy subjects with minimum scores on the SIS-R, among whom no further distinction can be made. Since the CAPE was specifically designed to assess symptoms in low-scoring individuals, it would be an important advantage for both research and clinical work if the information contained in CAPE items could be somehow used to improve the precision of the estimation of subclinical psychotic symptoms based on the SIS-R. Here we will use CAPE items to enhance measurement precision of the SIS-R scores by modeling two correlated latent traits, one for CAPE items and one for SIS-R items, through a MIRT model.

Before we combine the information from the CAPE and SIS-R, we will investigate the dimensionality of the instruments separately using three complementary methods: Mokken Scale Analysis (MSA), multidimensional Item Response Theory models (MIRT), and ordinal factor analysis (FA). In addition, we will test whether the assessment of schizotypy is influenced by individual characteristics, such as being a sibling of a schizophrenia patient. It is not unlikely that siblings interpret items differently compared to community controls, as they have been in close personal contact with a psychotic family member: they probably have better knowledge of what might be involved regarding certain symptom descriptions. As a consequence, the item score of a given person may depend not only on the latent dimensions of interest but will also depend on individual characteristics (Mellenbergh, 1989; Meredith, 1993). Such a violation of measurement invariance complicates a fair comparison of liability scores across groups.

2. Methods

2.1. Subjects

The data were collected as part of the Genetic Risk and Outcome of Psychosis (GROUP) project (www.group-project.nl), a longitudinal observational study focusing on the factors that make people vulnerable to develop psychosis (GROUP, 2011). Eligible siblings of schizophrenia patients had to fulfill the criteria of (1) age between 18 and 50 (extremes included), (2) fluent in Dutch, and (3) able and willing to give written informed consent. Eligible healthy controls had to fulfill the criteria of (1) age between 18 and 50 (extremes included), (2) no lifetime psychotic disorder, (3) no first-degree family member with a lifetime psychotic disorder, (4) fluent in Dutch, and (5) able and willing to give written informed consent. In the present study we included a sample of 1088 subjects (639 siblings of schizophrenia patients and 327 controls with CAPE data; 746 siblings and 339 controls with SIS data) who had been assessed at the research center in Utrecht, Groningen, or Amsterdam. The mean age of controls was 31 years (S.D.=10.5; 41.5% male) and the mean age of the siblings was 27 years (S.D.=8.0; 46.3% male).

2.2. Measures

The Dutch versions of the Community Assessment of Psychic Experiences (CAPE) and The Revised Structure Interview for Schizotypy (SIS-R) were assessed. The CAPE is a self-report tool measuring lifetime subthreshold psychotic experiences. It consists of 42 items assessing the frequency (rated on a 4-point Likert scale) of subclinical psychotic experiences in the following three domains: positive symptoms (20 items), negative symptoms (14 items) and depression symptoms (8 items).

The SIS-R (Kendler et al., 1989; Vollema and Ormel, 2000) is an interview instrument that measures a broad range of schizotypal symptoms and signs by applying standardized rating and scoring procedures (four response categories). The shortened version of the SIS-R used in this study describes schizotypy in two dimensions: positive schizotypy (7 items) and negative schizotypy (8 items). It should be noted that we consider both the CAPE and SIS-R to be indicators of schizotypy, even though the CAPE refers to the measured construct as ''subclinical psychosis''; both measures include subscales tapping into both positive and negative symptoms.

2.3. Statistical analyses

2.3.1. Assessing dimensionality of CAPE and SIS-R

Three complementary techniques were used to investigate the dimensionality of the CAPE and SIS-R: Mokken Scale Analysis, parametric IRT analysis, and ordinal factor analysis. Mokken Scale Analysis (MSA; Mokken, 1971; Sijtsma et al., 2011) was applied using the software package Mokken Scale Analysis for Polytomous items (MSP5.0; Molenaar and Sijtsma, 2000). MSA is a non-parametric type of IRT analysis. MSA can be used to uncover the dimensionality (factorial structure) of the data, and at the same time identifies scales that allow an ordering of individuals on an underlying one-dimensional scale using the unweighted sum of item scores. In order to determine which items belong together and form a scale, scalability coefficients are calculated. Similar to the item-rest correlation, the scalability coefficient expresses the degree to which an item is related to other items in the scale. The scalability coefficient can be seen as a 'corrected' correlation: the correlation between items is divided by the maximum expected correlation given the items' marginal score-frequency distributions. Dimensionality was investigated using MSP5.0's automated item selection procedure (AISP) that aims to find one-dimensional clusters of items. These clusters were identified by running the AISP several times in a row, each time increasing the lower bound scalability coefficient (also known as the user-specified constant, c). Following (Sijtsma and Molenaar, 2002; see also Meijer et al., 2011), we ran the AISP repeatedly for increasing values of c. The resulting sequence of outcomes indicates whether the data set is one-dimensional or multidimensional. Sijtsma and Molenaar (2002) provide the following guidelines. In case of one unidimensional scale for all items, the typical sequence is (1) most or all items are in one scale, (2) one smaller scale is found, and (3) one or a few small scales are found and several items are excluded. In multidimensional datasets the typical sequence is (1) most or all items are in one scale, (2) two or more scales are formed, and (3) two or more smaller scales are formed and several items are excluded. For a recent empirical application of this procedure see Wismeijer (2012).

Parametric IRT models have the same basic assumptions as Mokken models: one-dimensionality, monotonicity and local independence (Reise and Waller, 2009). The difference is that where the Mokken scale merely assumes a non-decreasing relation between the probability of a positive response as a function of trait level, IRT models assume a parametric form for this relationship, either a logistic function or a normal probability distribution, so that IRT models are more restrictive than Mokken models, but allow for the possibility that some items are better indicators for a trait than others. The specific model used here was the Generalized Partial Credit Model (GPCM, Muraki, 1992) for polytomous items. Moreover, we applied multidimensional extensions of the GPCM, where we assumed that individuals have two or more latent trait levels, which might be correlated. Each latent trait is coupled to a fixed set of items, for instance the positive or the negative symptom items on the SIS-R, so that each latent trait can be interpreted through the items associated with it (Béguin and Glas, 2001). Marginal Maximum Likelihood estimation was used. Model fit was ascertained by computing absolute differences between expected and observed item scores for high, average and low scoring individuals. An absolute difference smaller than 0.10 was interpreted as sufficient item fit (cf. Van den Berg et al., 2010). The parametric IRT analyses were applied using the package MIRT (Glas, 2010).