Acta Neuropsychiatrica 2018
All rights reserved
DOI: 10.1017/neu.2017.40

© Scandinavian College of Neuropsychopharmacology 2018
ACTA NEUROPSYCHIATRICA

A clinimetric analysis of a BPRS-6 scale for schizophrenia severity

Bech P, Austin SF, Timmerby N, Ban TA, Møller SB. A clinimetric analysis of a BPRS-6 scale for schizophrenia severity.

Objective: A restricted Brief Psychiatric Rating Scale (BPRS-6) with the six schizophrenia specific items from the Positive and Negative Syndrome Scale (PANSS) has been investigated. These six items from the PANSS have recently been found to have both clinical validity and 'unidimensionality' in measuring the severity of schizophrenic states. The primary objective of this study was to evaluate the clinical validity of the BPRS-6. The secondary objective was to evaluate the 'unidimensionality' of the BPRS-6 by an 'item response theory' model.
Methods: The BPRS-6 was scored independently by two psychiatrists and two psychologists while viewing six open-ended videotaped interviews in patients with a DSM-III diagnosis of schizophrenia. The interviews were conducted by Heinz E. Lehmann, an experienced psychiatrist. They were focused on the psychopathology that contributed most to the 'severity' of the patient's clinical state.
Results: The BPRS-6 with three positive symptoms (delusions, conceptual disorganisation, hallucinations) and three negative symptoms (blunted affect, emotional withdrawal, poverty of speech) was found to be clinically valid and captured the variables that contribute most to the severity of schizophrenia. The BPRS-6 was also found to have acceptable 'unidimensionality' (coefficient of homogeneity 0.45) and inter-rater reliability (inter-class-coefficient 0.81).
Conclusion: The BPRS-6 was found to capture the information that translates into the severity of schizophrenia. It has also acceptable psychometric validity.

Per Bech[1], Stephen F Austin[1], Nina Timmerby[1], Thomas A Ban[2], Stine Bjerrum Møller[1]

[1]Psychiatric Research Unit, Psychiatric Centre North Zealand, Copenhagen University Hospital, Hillerød, Denmark; and [2]Vanderbilt University, Nashville, Tennessee, USA

Keywords: antipsychotic; psychiatric disorders; schizophrenia

Per Bech, Psychiatric Research Unit, Mental Health Centre North Zealand, University of Copenhagen, Dyrehavevej 48, DK-3400 Hillerød, Denmark.
Tel: +45 38 64 30 95;
Fax: +45 38 64 30 99;
E-mail: Per.bech@regionh.dk

Accepted for publication December 19, 2017

First published online February 7, 2018

Significant outcome

- The Brief Psychiatric Rating Scale 6 (BPRS-6) was found to have both clinical and psychometric validity in measuring the severity of schizophrenia.

Limitation

- The study was limited to the evaluation of six patients with a DSM-III diagnosis of schizophrenia by four experienced clinicians.
- It was a cross-sectional study without an evaluation of test-retest reliability and sensitivity to measure change during treatment.

Introduction

Rating scales such as the Hamilton Depression Scale (HAM-D) (1), the BPRS (1) or the Bech Rafaelsen Mania Scale (MAS) (1) were developed to measure the changes in the clinical state resulting from treatment with antidepressant, antipsychotic or

https://doi.org/10.1017/neu.2017.40 Published online by Cambridge University Press

anti-manic drugs respectively (2). The first step in a clinimetric analysis of these scales includes an evaluation of their content or face validity performed by experienced psychiatrists (3).

From a measurement point of view, the second and probably most important step in a clinimetric analysis of rating scales is the testing of their construct validity. In clinimetrics or clinical psychometrics (4) the evaluation of construct validity is performed by using item response theory (IRT) models introduced in the testing of psychiatric rating scales by Bech et al. (5,6).

The first IRT model was developed by Rasch (7). It entails the testing of each item of a scale to detect to what extent it stands alone on the dimension of severity tested by the scale. By mapping the items in three compartments, one can identify the items covering the low end of the dimension (contribute least to the determination of severity), the items covering the severe end (contribute most to the determination of severity), and the items covering the part of the dimension in between (1).

For the Hamilton Depression Scale (HAM-D), we have shown that only six of the 17 items in this scale (8) have clinimetric validity (2,6,9). Concerning the BPRS, its clinimetric validity has been insufficiently evaluated (1). Yet, when developing the Positive and Negative Syndrome Scale (PANSS) Kay (10) used the BPRS as his orientation point for selecting the 30 items (PANSS-30) of the scale.

In the PANSS-30 there is the same number of items (namely seven) for positive and for negative symptoms. In the BPRS, however, Kay (10) identified eleven positive symptoms and only three negative symptoms.

A brief scale for rating chronic psychotic states overlapping with the core items in BPRS was developed by Krawiecka et al. (11). In this scale positive symptoms, such as hallucinations, conceptual disorganisation, and delusions and negative symptoms such as poverty of speech, blunted affect, and emotional withdrawal were included.

We have recently evaluated the PANSS-30 (12) using the parametric IRT Rasch model. We found that the full PANSS-30 had no scalability in contrast to a PANSS-6 subscale which according to Andreasen et al. (13) includes items with content validity for schizophrenia. Out of these six items, three measured negative symptoms (lack of spontaneity, blunted affect and socials withdrawal) and three measured positive symptoms (delusions, conceptual disorganisations and hallucinations). However, as indicated in Fig. 1 the item of social withdrawal in PANSS-6 is based on reports from care-workers or family, whereas all six items in the BPRS-6 are based on observation at interview.

The BPRS items were originally graded by Overall and Gorham (14) as the PANSS items on a Likert scale from 1 (not present) to 7 (present in extreme degree) but, Overall and Hollister (15) changed the Likert scale from 0 (not present) to 6 (present in extreme degree) because it is statistically much more appropriate to have 0 as 'not present' consistent with the majority of rating scales used in psychiatry.

The PANSS-30 was developed in its clinician-reported version in four steps: (1) open-ended interview; (2) semi-structured interview; (3) structured interview; and (4) 'directive interview' (10). It was the structured – directive interviews of the PANSS-30 that stimulated several pharmaceutical firms to use the PANSS in multi-centre studies of new drugs for schizophrenia (10). Furthermore, it was the inter-rater reliability of non-experienced clinicians of these interviews that was the advantage of PANSS when compared with the other rating scales used in this field, for example the BPRS.

According to Kay (10), the construct validity of the PANSS-30 was tested only by factor analysis, a method for identifying structures, for example negative versus positive factors, but not for testing the measurement dimension of severity in schizophrenia.

The first clinimetric validity study of the PANSS-30 showed that only six of the items (Fig. 1) had both content (or face) validity and construct validity (12).

In the present study, videotaped interviews conducted by Professor Heinz E Lehmann, an experienced psychiatrist (16), were used to score the BPRS. The patients participating in the interviews fulfilled the DSM-III diagnosis of schizophrenia. Each videotape was prepared for teaching clinical investigators how to use the BPRS.

The interviews were open, non-structured interviews which complied with the rules of a clinimetrically valid interview as stipulated by Roth (17).

The objective of our study was to evaluate the extent to which the BPRS items in Fig. 1 were captured by Lehmann in his interviews (content validity), and to determine the construct validity of these six items as tested by the use of the non-parametric IRT analysis of Mokken (18). Inter-rater reliability of the BPRS-6 was also evaluated.

Method

The patients

The interviews performed by Heinz Lehmann were carried out in 1982 and all six patients fulfilled DSM-III criteria of schizophrenia. Four patients were females: 28, 29, 38 and 35 years of age. Two patients were males: 24 and 26 years of age.

https://doi.org/10.1017/neu.2017.40 Published online by Cambridge University Press

PANSS-6			BPRS-6		
P1	Delusions (at interview)	(1-7)	1	Expressed delusions (at interview)	(0-6)
P2	Conceptual disorganization (at interview)	(1-7)	2	Conceptual disorganization (at interview)	(0-6)
P3	Hallucinations (at interview)	(1-7)	3	Hallucinations (at interview)	(0-6)
N1	Blunted affect (at interview)	(1-7)	4	Blunted, incongruous affect (at interview)	(0-6)
N4	Passivity - Social withdrawal (reports from care-workers or family)	(1-7)	5	Emotional withdrawal (at interview)	(0-6)
N6	Lack of spontaneity and flow of conversation (at interview)	(1-7)	6	Poverty of speech (at interview)	(0-6)
	Theoretical score range	(6-42)		Theoretical score range	(0-36)

Fig. 1. The scoring sheet of the Positive and Negative Syndrome Scale 6 (PANSS-6) and Brief Psychiatric Rating Scale 6 (BPRS-6).

The BPRS interviews

All six interviews were open interviews, beginning with questions concerning sleep and appetite in order to facilitate a dialogue. The duration of each interview was ~25 min. Not all BPRS items were evaluated, but all six items shown in Fig. 1 were assessed, as Lehmann considered them central in the psychopathology of schizophrenia, and, as such, they were the target symptoms for antipsychotic therapy.

These interviews were conducted at the Douglas Hall of Douglas Hospital, one of the teaching hospitals of McGill University, in Montreal, Canada and are available at the CAMH Archives, in Toronto, Canada (16).

Each of the six items of the modified BPRS presented in Fig. 1 was scored on a global Likert scale: 0 = not present; 1 = doubtfully present; 2 = mild; 3 = moderate; 4 = marked; 5 = severe; 6 = extremely severe.

The raters

All four raters were experienced psychiatrists (P.B. and N.T.) and clinical psychologists (S.B.M. and S.A.), who scored each interview independently.

Statistical analysis

Inter-rater reliability

To evaluate inter-rater reliability the intra-class coefficient was used (19,20).

Scalability

To evaluate scalability, a programme of the Mokken non-parametric IRT model (18), developed by Molenaar et al. (21), was used. In this model, the coefficient of homogeneity (or scalability) expresses to what extent each item of the scale contributes to the measurement of the severity of schizophrenia from its lowest to its highest extremes. With the employment of this method, a coefficient of 'scalability' between 0.30 and 0.39 indicates that the scale is just marginally acceptable for measuring scalability, whereas a coefficient of 0.40 or higher indicates that the scale is definitely acceptable to measure it (18).

Results

Inter-rater reliability evaluation

Table 1 shows the total BPRS-6 scores for the four raters on the six patients. The intra-class coefficient of raters was 0.81, that is an acceptable inter-observer reliability (20).

The scalability evaluation

The Mokken coefficient of homogeneity, or coefficient of scalability, was 0.45 for the total summed item score of BPRS-6. For the individual items the coefficient of homogeneity was for item 1: 0.45; for item 2: 0.50; for item 3: 0.10; for item 4: 0.56; for item 5: 0.55; and for item 6: 0.56. This indicates that the total score of the 6 BPRS items (Table 1) is a sufficient statistic for the measurement of the severity of schizophrenia.

Discussion

The construct validity of the BPRS-6 scale used in the present study was tested by the non-parametric Mokken model. The Mokken analysis was preferred to the Rasch analysis is this study because of the very small number of patients evaluated.

https://doi.org/10.1017/neu.2017.40 Published online by Cambridge University Press

Table 1. Total Brief Psychiatric Rating Scale 6 score for each of the four raters for the six patients.

	The four raters			
The six interviewed patients	PB	NT	SBM	SA
A	13	15	17	14
B	29	33	30	29
C	19	22	22	13
D	26	24	22	24
E	8	11	14	14
F	30	30	25	19

A major limitation of this study is that we only had access to the six videotaped interviews. In such a case with a small sample size ($n = 6$) there is no alternative to using a non-parametric test (22). This small number of observations implies that the results have to be very consistent. The non-parametric Mokken coefficient of homogeneity was for the full BPRS-6 0.45 indicating that the summed scale score is a sufficient statistic (18). The inter-rater reliability of 0.81 is also acceptable (20).

A further limitation of the study is that we have not been able to compare the BPRS-6 with other schizophrenia rating scales in evaluating such psychometric issues as convergent versus discriminant validity. As we only had access to the baseline ratings when the patients had been hospitalised we have not been able to evaluate the test–retest reliability versus the sensitivity to measure change in the course of treatment.

The PANSS-30 was developed in its final version for use by inexperienced clinicians often selected to participate in industry-sponsored multi-centre studies on new drugs in the treatment of schizophrenia (10). However, the BPRS was developed by Overall and Gorham (14) to be administrated in a flexible type of interview to increase the clinical validity of the scale (15).

Tomba and Bech (4) argue that the initial clinical judgment of experienced psychiatrists preferred by Feinstein (3) needs to be complemented by a detailed assessment of the symptoms using unidimensional scales in order to perform a rich and valid assessment of the patient.

The six BPRS interviews performed by the experienced psychiatrist Heinz Lehmann illustrates how his clinical skill, intuition and insight during the interviews have captured the six items shown in Fig. 1 as the target symptoms of schizophrenia.

In DSM-5, all six items in Fig. 1 are included in the symptomatic criteria for schizophrenia. The three items at the top in Fig. 1, expressed delusions, conceptual disorganisation and hallucinations, provide the items for the diagnostic criteria of psychosis. The three other items in Fig. 1, blunted incongruous affect, emotional withdrawal and poverty of speech, provide for the diagnostic criteria of schizophrenia.

Recovery can be expressed either as a categorical variable (present/absent) or as a comparative category (non-recovered, slightly recovered, moderately recovered and greatly recovered) (4). Clinimetrics is useful in capturing the dimensional perspective of recovery. Sensitivity towards subtle improvements of the psychotic condition could affect treatment decisions and the spirit of the patient. We agree with Katz (23) that the definition of schizophrenia in DSM-5 is one of the few attempts to use a dimensional approach to mental disorders.

In conclusion, the BPRS-6 scale shows promise as a flexible unidimensional rating scale to measure the severity of schizophrenia. This study can be considered as a confirmation of the clinimetric validity of the PANSS-6. When used by experienced clinicians, as recommended by Lehmann et al. (24), an acceptable inter-rater reliability of the BPRS-6 was obtained.

In measurement-based care, time has come to focus on the determination of the clinimetric validity of rating scales evaluating the results of the various interventions used in the treatment of mental disorders (2).

Both DSM-5 and the industry-initiated clinical trials with psychotherapeutic drugs point to the need for inter-rater reliability. However, adequate inter-rater reliability is an integral part of clinimetrically validated scales (25).

Acknowledgements

Authors' Contributions: all authors contributed equally and fulfil the International Committee of Medical Journal Editors (ICMJE) authorship criteria.

Financial Support

The research received no specific grant from any funding agency, commercial, or non-profit-sectors.

Conflicts of Interest

The authors declare there are no conflicts of interest pertaining to this work.

Ethical Standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

References

1. Bech P. Clinical psychometrics. Oxford: Wiley Blackwell, 2012.

https://doi.org/10.1017/neu.2017.40 Published online by Cambridge University Press

2. Bech P. Measurement-based care in mental disorders. New York: Springer, 2016.
3. Feinstein AR. Clinimetrics. New Haven: Yale University Press, 1987.
4. Tomba E, Bech P. Clinimetrics and clinical psychometrics: macro- and microanalysis. Psychother Psychosom 2012; **81**:333–343.
5. Bech P, Allerup P, Rosenberg R. The Marke-Nyman temperament scale. Evaluation of transferability using the Rasch item analysis. Acta Psychiatr Scand 1978;**57**:49–58.
6. Bech P, Allerup P, Gram LF et al. The Hamilton depression scale. Evaluation of objectivity using logistic models. Acta Psychiatr Scand 1981;**63**:290–299.
7. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research. Expanded Edition Chicago: Chicago University Press 1980, 1960.
8. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967;**6**:278–296.
9. Timmerby N, Andersen JH, Søndergaard S, Østergaard SD, Bech P. A systematic review of the clinimetric properties of the 6-item version of the Hamilton Depression Rating Scale ($HAM\text{-}D_6$). Psychother Psychosom 2017;**86**:141–149.
10. Kay SR. Positive-negative symptom assessment in schizophrenia: psychometric issues and scale comparison. Psychiatr Q 1990;**61**:163–178.
11. Krawiecka M, Goldberg D, Vaughan M. A standardized psychiatric assessment scale for rating chronic psychotic patients. Acta Psychiatr Scand 1977;**55**:299–308.
12. Ostergaard SD, Lemming OM, Mors O, Correll CU, Bech P. PANSS-6: a brief rating scale for the measurement of severity in schizophrenia. Acta Psychiatr Scand 2016;**133**:436–444.
13. Andreasen NC, Carpenter WT Jr, Kane JM, Lasser RA, Marder SR, Weinberger DR. Remission in schizophrenia: proposed criteria and rationale for consensus. Am J Psychiatry 2005;**162**:441–449.
14. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. Psychological reports 1962;**10**:799–812.
15. Overall JE, Hollister LE. Assessment of depression using the Brief Psychiatric Rating Scale. In Sartorius N, Ban TA editors Assessment of depression. Heidelberg: Springer-Verlag, 1986; p. 159–178.
16. Blackwell B. Heinz Edgar Lehmann, 2015. Available at inhn.org. Accessed 19 April 2016.
17. Roth M. The clinical interview and psychiatric diagnosis: have they a future in psychiatric practice? Compr Psychiatry 1967;**8**:427–428.
18. Mokken RJ. Theory and practice of scale analysis. Berlin: Mouton, 1971.
19. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. Psychol Rep 1966;**19**:3–11.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;**33**:159–174.
21. Molenaar IW, Debels P, Sijtsna K. User's manual MSP, a program for Mokken scale analyses for polytomous items (version 3.0). Groeningen, the Netherlands: ProGAMMA, 1994.
22. Siegel S. Nonparametric statistics for the behavioural sciences. New York: McGraw Hill, 1956.
23. Katz MM. Clinical trials of antidepressants. New York: Springer Verlag, 2016.
24. Lehmann HE, Ban TA, Donald M. Rating the rater. An experimental approach to the methodological problem of interrater agreement. Arch Gen Psychiatry 1965;**13**:67–75.
25. Bech P. Rating scales for psychopathology, health status and quality of life. A compendium on documentation in accordance with the DSM-III-R and WHO systems. Berlin: Springer, 1993.

https://doi.org/10.1017/neu.2017.40 Published online by Cambridge University Press