

COSMIN systematic review and meta-analysis of the measurement properties of the Positive and Negative Syndrome Scale (PANSS)



Simon Geck,^{a,g} Maximilian Roithmeier,^{a,g} Markus Bühner,^{b,c} Sophia Wehr,^a Lucia Weigel,^a Josef Priller,^{a,c,d,e} John M. Davis,^f and Stefan Leucht^{a,c,*}



^aDepartment of Psychiatry and Psychotherapy, School of Medicine and Health, Technical University of Munich, Klinikum Rechts der Isar, Ismaningerstraße 22, Munich, 81675, Germany

^bDepartment of Psychology, Ludwig Maximilian University of Munich, Leopoldstr. 13, Munich, 80802, Germany

^cGerman Center for Mental Health (DZPG), Site Munich/Augsburg, Germany

^dNeuropsychiatry and Laboratory of Molecular Psychiatry, Charité – Universitätsmedizin Berlin and DZNE, Berlin, Germany

^eCentre for Clinical Brain Sciences, UK Dementia Research Institute at the University of Edinburgh, Edinburgh, UK

^fPsychiatric Institute, University of Illinois at Chicago (mc 912), 1601 W. Taylor St., Chicago, IL, 60612, USA

Summary

Background The Positive and Negative Syndrome Scale (PANSS) is the most widely used tool for assessing the symptoms of schizophrenia. Despite its widespread use, the psychometric properties of the PANSS have not been systematically reviewed. This study fills that gap in the scientific literature.

Methods We utilized the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guideline for systematic reviews and meta-analytical procedures to assess the psychometric properties of the PANSS in its original three-subscale form as well as the quality level of the evidence available. On this basis we formulated recommendations for future research and use. A study protocol was registered under 10.17605/OSF.IO/5EGMD. The search period was until February 21, 2024.

Findings We included 119 publications. According to COSMIN, the PANSS demonstrated sufficient reliability, construct validity, and responsiveness; but had significant shortcomings in content validity and structural validity. The original three-factor model showed poor structural validity, leading to its COSMIN classification as “not recommendable”. The subscales showed overall acceptable measurement properties. However, the lack of structural validity of the three-subscale model renders its subscales less useful. Moreover, the PANSS negative subscale does not cover all domains of the National Institute of Mental Health consensus. Due to the length of the instrument (30–50 min), it is barely useable in clinical practice.

Interpretation Although the PANSS is the standard scale for schizophrenia symptom severity, its shortcomings regarding fundamental psychometric domains and practical applicability warrant the development of new scales for which appropriate methods should be applied from the start.

Funding There was no specific funding source for this research.

Copyright © 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Schizophrenia; Antipsychotic drug trials; Positive and Negative Syndrome Scale; PANSS; Psychometrics; Psychotic disorders; Rating scales; Psychopathology

Introduction

The Positive and Negative Syndrome Scale (PANSS) is the most frequently used scale to rate the symptoms of schizophrenia.¹ It was developed in the 1980s by Kay et al. when it became clear that negative symptoms

are at least as significant, if not more significant, than the positive symptoms of schizophrenia.^{1,2} For example, Crow had hypothesized that there is a type I schizophrenia characterized by positive symptoms and good response to antipsychotic drugs, and a type II

*Corresponding author. Technical University of Munich, School of Medicine and Health, Department of Psychiatry and Psychotherapy, Klinikum rechts der Isar, Ismaningerstraße 22, 81675 Munich, Germany.

E-mail address: stefan.leucht@tum.de (S. Leucht).

^gMr. Geck and Roithmeier contributed equally to this work and are to be recognized as first authors.

Research in context

Evidence before this study

The Positive and Negative Syndrome Scale (PANSS) is probably the most frequently used rating scale for assessing the symptoms of schizophrenia. Therefore, it is necessary to systematically review the measurement properties of the PANSS.

We therefore performed a search for reviews on the psychometric properties of the PANSS via PubMed. In 2015, Garcia-Portilla et al. conducted a systematic review on the measurement properties of the PANSS negative subscale but addressed neither the total PANSS nor the positive and general psychopathology subscale and further did not follow a structured and comprehensive approach. To date, no systematic review covering the full range of available data regarding the measurement properties of the whole PANSS scale and its three subscales was published. We fill this gap by applying the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines for systematic reviews of patient-reported outcome measures.

Added value of this study

We present the first systematic review and meta-analysis of the psychometric properties of the PANSS. Without language

restrictions, we searched the databases MEDLINE and Embase for evaluation studies of the PANSS published between the initial publication of the PANSS and February 2024 and identified 119 publications containing the assessment of 10 different psychometric properties. Quantitative results were meta-analytically pooled and no indications for small-trial bias have been found. We found that the PANSS has strengths in terms of reliability, construct validity, and responsiveness. There are weaknesses in the domains content validity and structural validity which are considered fundamental by COSMIN. The feasibility of the scale in practice is questionable. Research suggests a 5-factor structure rather than the original 3 subscale form.

Implications of all the available evidence

The PANSS has been successfully applied to separate between interventions in schizophrenia trials. Nevertheless, given its shortcomings in terms of content validity, structural validity, lack of covering important domains of negative symptoms and feasibility, the development with modern methods of a new scale to rate symptoms of schizophrenia is warranted. Moreover, we recommend further research into five-factor solutions of the PANSS and short versions.

schizophrenia predominated by negative symptoms and associated with poorer outcome.^{3,4} Similarly, Andreasen coined the terms positive and negative schizophrenia.² In practice the acute phase is often dominated by positive symptoms, while negative symptoms develop later, and some of these negative symptoms may be the result of extrapyramidal side-effects. At that time, the most commonly used 18-item version of the Brief Psychiatric Rating Scale (BPRS) did not sufficiently cover negative symptoms.² Therefore, a new scale had to be developed. Following the two-dimensional positive versus negative symptoms illness concept and adding a third domain for measuring general psychopathology, the PANSS was created.^{2,5} It is a synthesis of all 18 items of the BPRS with 12 items from the 32-item Psychopathology Rating Schedule (PRS), which Kay and colleagues had put together in the 1970s without, however, following valid scale-development methods.^{6,7} In order to improve on the BPRS, the severity levels of each of the 30 PANSS items were anchored by clear descriptions.¹ Thus, the PANSS consists of 3 subscales (positive scale, negative scale, and general psychopathology scale) with 7 items each on the positive and the negative subscale, and 16 items on the general psychopathology subscale, all rated between 1 “absent” and 7 “extreme”.¹

Despite the predominance of the PANSS in rating symptoms of schizophrenia, its psychometric properties have, to the best of our knowledge, never been addressed by a systematic review and meta-analysis. We fill this gap by applying the COnsensus-based Standards

for the selection of health Measurement INstruments (COSMIN) guideline, which was developed in an international Delphi study to evaluate the measurement properties of health-related patient-reported outcomes.^{8–10} In the current publication, we addressed the total score and the original three PANSS subscales. In companion publications, we review various factor solutions (e.g., the 5-factor solution found by Marder et al. 1997), and short-versions such as PANSS-6.^{11,12}

Methods

The methods used in this review are based on the COSMIN guideline for systematic reviews of patient-reported outcome-measures.^{8–10} COSMIN was initially created for patient-reported outcome-measures (PROMs) but can also be applied to clinician-reported outcome-measures (ClinROMs), such as the PANSS.^{13,14} There are several steps: *literature search*, quality assessment of the individual studies with the COSMIN *Risk-of-Bias checklist*, applying *updated criteria for good measurement properties*, *summarizing the evidence*, *grading the overall quality of evidence* with a modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, formulating a *recommendation* and describing *feasibility aspects*.^{8–10} An overview is presented in [Fig. 1](#).

All steps of the COSMIN methodology were conducted independently by two reviewers (SG and MR). Any disagreements were resolved by consensus, with the involvement of a third reviewer (SL, MB), or the COSMIN group (WM) if necessary.

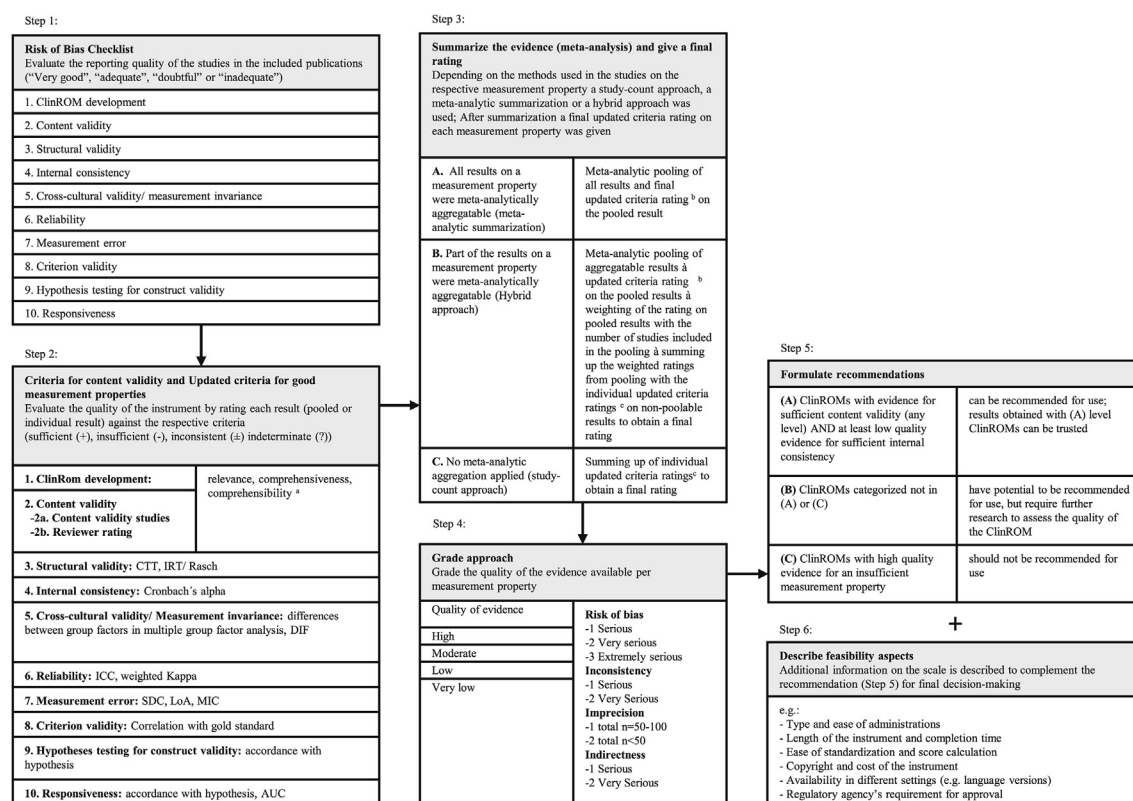


Fig. 1: Overview of the COSMIN methodology including adaptations. Refer to the methods part and appendix of our publication for further information. For the unmodified, full version of the COSMIN methodology, refer to the COSMIN manual. Step 1: Risk of Bias Checklist. Step 2: Criteria for content validity and Updated criteria for good measurement properties. ^aFor the detailed criteria refer to the COSMIN manual for content validity. Step 3: Summarize the evidence and give a final rating. ^bThe same criteria like in Step 2 apply. ^cThese are the individual ratings on each result obtained in Step 2. Step 4: Grade approach. Step 5: Formulate recommendations. Step 6: Describe feasibility aspects.

A protocol was registered on the Open-Science-Foundation website where more details can be found (<https://doi.org/10.17605/OSF.IO/5EGMD>; compare Appendix S1).¹⁵

Search strategy and selection criteria

A PhD librarian (FS) searched PubMed and EMBASE using a modified COSMIN filter, see Appendix S2 and Terwee et al. (2009), for journal articles published until February 21, 2024.¹⁶ We screened the results and reference lists of included publications for evaluation studies of the PANSS in which at least 80% of study participants had a psychotic disorder in any stage (e.g., acute or in remission). There were no language or country restrictions. Following COSMIN, only full-text articles were included.^{8–10} We addressed the original PANSS total, positive, negative and general scores.¹

Assessing the risk of bias (step 1 in Fig. 1)

The COSMIN risk of bias checklist consists of 10 domains (called “boxes” by COSMIN): (1) ClinROM development, (2) content validity, (3) structural validity,

(4) internal consistency, (5) cross-cultural validity/measurement invariance, (6) reliability, (7) measurement error, (8) criterion validity, (9) hypothesis testing for construct validity, (10) responsiveness.^{8–10} Each box consists of two to eight items rated as either “very good”, “adequate”, “doubtful”, “inadequate”, or “not applicable”. A worst score counts principle is applied for each domain.

For criterion validity, the Brief Psychiatric Rating Scale (BPRS) was considered as the gold-standard because of its reported empirical development, stable factor structure, and good reliability.^{6,17} Risk of Bias (RoB) within COSMIN addresses whether and how well these domains were assessed. The results (e.g., whether Cronbach's alpha was acceptably high) are part of the next step.

Assessing the criteria for content validity and the updated criteria for good measurement properties (step 2 in Fig. 1)

The results of each study are evaluated following COSMIN's criteria for I. content validity (Fig. 1, step 2,

domains 1 and 2) and II. The *updated criteria for good measurement properties*.^{8–10} (Fig. 1, step 2, domains 3–10).

A study's evidence for each domain is rated as either "sufficient" (+) or "insufficient" (–), with a third option for an "indeterminate" rating (?), if neither "sufficient" nor "insufficient" ratings are applicable. Only for content validity (domains 1 and 2), a fourth rating option of "inconsistent" (±) is available.

The overall rating for I. *Content validity* is generally made up of three individual ratings for the ClinROM development study, content validity studies, and a reviewer rating of the scale (see Appendix S3).

The II. *Updated criteria* address domains 3–10. Concerning domain 3 "Structural validity", the COSMIN guideline does not provide criteria for factor models resulting from exploratory methods. After consultation with the COSMIN group, we therefore created criteria based on a modified version of the *Criteria for good measurement properties* used by Elsmann et al. (2022), for the purpose of comparing the exploratory derived models to Kays' three-factor model.¹⁸ The criteria are: 1) ≥50% explained variance, 2) factor loadings >0.30, 3) ≤10% cross-loading items, 4) three-factor model, 5) ≥80% of the items distributed like in the Kay three-factor model. A "sufficient" rating was only given when all of the criteria regarding the exploratory derived models were met. Otherwise, the respective rating was "insufficient". The criteria for structural validity established by Elsmann et al. (2022) are identical for both exploratory factor analyses (EFAs) and principal component analyses (PCAs); hence, we generalized this approach to the entire review.¹⁸

For assessing domains 9 "construct validity" and 10 "responsiveness", COSMIN requires a formulation of hypotheses which are presented in Appendix S4.^{8–10}

We considered correlations of the PANSS total score and the subscores with other scales; correlations of single PANSS items were not considered.

For details about the assessment of all other domains suggested by COSMIN, please see pp.28 and 29 of the COSMIN manual.^{8–10}

If uncertainties arose regarding results or methods of relevant articles, their authors were contacted and asked to revalidate the results or processes. This applied to 39 articles.

Statistical method

In situations where few studies are available, a simple study-count approach is sufficient in COSMIN reviews, see e.g., Weigel et al. (2023), Wehr et al. (2024) or Zúñiga Le-Bert et al. (2024).^{13,14,19} As our review includes 119 studies, we were able to summarize measures of internal consistency, cross-cultural validity, reliability, criterion validity, and construct validity with standard random-effects meta-analyses using Comprehensive Meta-Analysis (CMA) Version 2.²⁰ In situations where it was not possible to aggregate *all* studies, we used a hybrid of the study-count and the meta-analytic

approach, i.e., the meta-analytic result weighted by the number of studies included and summed up with ratings of studies that couldn't be aggregated and were therefore rated individually.

We measured heterogeneity with the I^2 -statistic and a chi-square test (compare Appendix S5); and small-trial/publication bias with funnel-plots and Egger's test when at least 10 studies were available. In most studies, several raters assessed multiple study participants. In some studies, multiple raters assessed small sample sizes ($n \leq 4$). CMA cannot process sample sizes $n \leq 4$.²⁰ We included the respective studies by exchanging the rater and patient count. Sensitivity analysis without them showed no important differences, allowing us to include these studies.

Grading the quality of evidence (step 4 in Fig. 1)

The COSMIN modified GRADE approach applies to the accumulated evidence on each measurement property, spanning 4 domains: (1) risk of bias, (2) inconsistency, (3) imprecision, and (4) indirectness.^{8–10}

The final quality of the evidence is graded as high, moderate, low, or very low (Table 1a). Initially, evidence is assumed to be of high quality. If the criteria of the domains are not fulfilled, the evidence is then downgraded by one to three levels (see Table 1b).

Recommendations (step 5 in Fig. 1)

According to COSMIN, an overall judgment of (A) "can be recommended for use and results obtained with the scale can be trusted", (B) "potential for recommendation but needs further research into the measurement properties", or (C) "should not be recommended for use" was made.^{8–10}

Recommendations are primarily dependent on the evidence for content validity and internal consistency (compare Fig. 1, Step 5).

Feasibility (step 6 in Fig. 1)

To supplement the recommendations, we compiled information on the feasibility of the PANSS (compare Fig. 1, Step 6).

Role of the funding source

There was no funding source for this study.

Results

Literature search

The search yielded 8070 results. Their titles and abstracts were assessed and 508 articles were selected for full-text screening. 110 publications were identified as eligible for this review. The additional screening of references resulted in 9 further inclusions. A flowchart depicting this process can be found in Fig. 2. For the general characteristics of included articles populations, refer to Appendix S6.

a) Quality of evidence definitions

High	We are very confident that the true measurement property lies close to that of the estimate of the measurement property
Moderate	We are moderately confident in the measurement property estimate: the true measurement property is likely to be close to the estimate of the measurement property, but there is a possibility that it is substantially different
Low	Our confidence in the measurement property estimate is limited: the true measurement property may be substantially different from the estimate of the measurement property
Very low	We have very little confidence in the measurement property estimate: the true measurement property is likely to be substantially different from the estimate of the measurement property

b) Instructions on downgrading**Risk of bias**

No	There are multiple studies of at least adequate quality, or there is one study of very good quality available
1 Serious	There are multiple studies of doubtful quality available, or there is only one study of adequate quality
-2 Very serious	There are multiple studies of inadequate quality, or there is only one study of doubtful quality available
-3 Extremely serious	There is only one study of inadequate quality available

Inconsistency

No	Results could be meta-analytically pooled or $\geq 75\%$ of the studies had consistent results
-1 Serious	$< 75\%$ of studies had consistent results
-2 Very serious	$< 70\%$ and $\geq 65\%$ had consistent results If $< 65\%$ had consistent results GRADE will not be performed because there is not enough evidence

Imprecision^a As recommended by COSMIN due to the high total sample sizes no downgrading for imprecision was done

Indirectness As recommended by COSMIN due to our inclusion criteria of $\geq 80\%$ population with a psychotic disorder no downgrading for indirectness was done

^aDowngrading for Imprecision is not applicable for measurement properties in which the sample size is already evaluated in the "Risk of bias" step (studies on content validity, structural validity and cross-cultural validity/measurement invariance).

Table 1: GRADE approach.

Assessing the risk of bias

Table 2 provides an overview of the distribution of the risk of bias and *Updated criteria* ratings.

ClinROM development and *content validity* of the PANSS could not be well assessed, because only the original study of the scale provided data and was rated as "doubtful". In the ClinROM development study of the PANSS, only a synthesis of two existing scales was done.¹ Thus, the concept elicitation and rational development of the PANSS were not executed in a sufficiently thorough manner.

A large number of studies assessing the *structural validity* of the PANSS total was identified.^{5,11,12,20–90} Most of them (60.90%) received "adequate" or "very good" ratings. Only few studies on the structural validity of the subscales were identified.^{49,91,92} All of them received "adequate" or "very good" ratings.

All studies on the internal consistency of the PANSS total received "inadequate" ratings, as internal consistency statistics should not be calculated for scores of scales that are not unidimensional or consist of subscores.^{21,29,37,74,80,93–98} Most studies (90.91%–93.75%) on the internal consistency of the subscales received "very good" ratings.^{1,55,74,80,82,98–108}

Only few studies on the *cross-cultural validity* of the PANSS total and the three subscales were found and all of them received either "doubtful" or "inadequate" ratings.^{109–112}

The majority (76.92%) of the studies on the *interrater reliability* of the PANSS total received "adequate" or

"very good" ratings.^{80,95,96,101,102,107,113–116} No studies on interrater reliability received "inadequate" ratings, but most of the studies (55.56%–58.82%) on the subscales received "doubtful" ratings.^{80,82,93,101,102,105–107,109,110,113–115,117}

Only few studies on *test-retest reliability* were found and all of them received "doubtful" or "inadequate" ratings.^{1,96,98,118}

Only few studies on *measurement error* were found for the PANSS total and its subscales, and all studies received "doubtful" or "inadequate" ratings.^{98,116}

Only few studies on the *criterion validity* of the PANSS total were found.^{101,102} All of them received "very good" ratings.

Only few studies for *hypothesis testing for construct validity* of the PANSS total were found and the majority (71.43%) of them received "doubtful" ratings.^{21,95,118–120} Most studies on construct validity were found for the positive and negative subscales. The majority (80.00%–81.58%) of the studies on the construct validity of the subscales received "adequate" or "very good" ratings.^{60,80,82,92,94–96,99–101,103,105,107,117,118,121–124}

Only few studies on the *responsiveness* of the PANSS total were found and none on the responsiveness of the subscales.^{29,120,125–127} All those studies received "very good" ratings.

More detailed information on the RoB rating process is available from the corresponding author upon request.

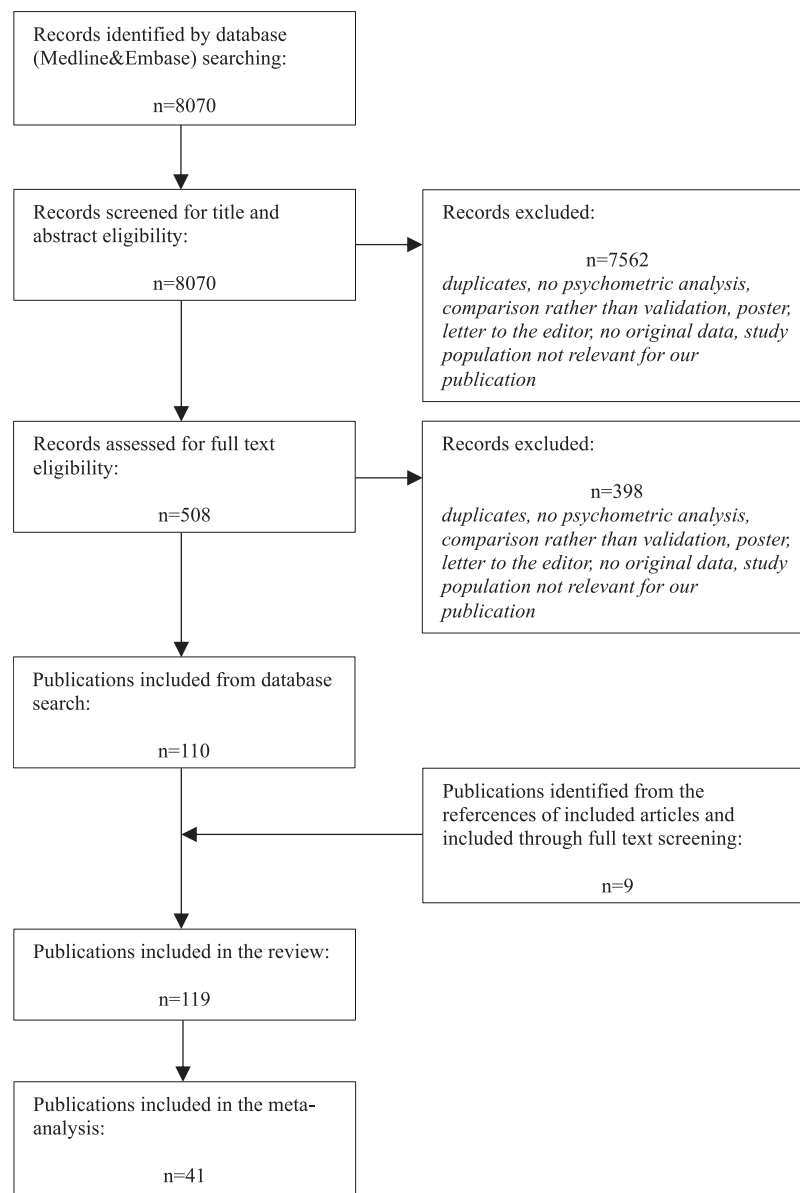


Fig. 2: PRISMA diagram of the search.

Assessing the criteria for content validity and the updated criteria for good measurement properties

Assessing the criteria for content validity

Due to the lack of content validity studies, the content validity rating is based on the rating of the ClinROM development study and the reviewer rating (Table 2).^{8–10}

The development study of the PANSS and the reviewer rating received “inconsistent” ratings due to weaknesses in *Relevance*, *Comprehensiveness* and *Comprehensibility*.¹ The overall content validity rating was “inconsistent”.

For a more comprehensive description of this rating process, we refer to Appendix S3.

Assessing the updated criteria for good measurement properties

These results are summarized in Table 2.

For the forest-plots and respective 95%-confidence intervals of the results from meta-analysis for the measurement properties *internal consistency*, *interrater-reliability*, and *construct validity*, we refer to Panel 1. For all remaining meta-analytical data, we refer to Appendix S7.

Internal structure.

Structural validity. 95 exploration-derived models and 16 confirmatory-factor-analyses of the original three-subscale

Measurement property	Scale (No. of studies assessing the scales measurement property)	COSMIN risk of bias				COSMIN updated criteria ratings ^a			
		Very good	Adequate	Doubtful	Inadequate	+ ^b	- ^c	? ^d	± ^e
ClinROM Development	PANSS total score (n = 1)	0	0	1	0	0	0	0	1
Content validity	PANSS total score (n = 0)	0	0	0	0	0	0	0	0
Structural validity	PANSS total score (n = 134)	28	53	24	28	0	111	22	
	Positive subscale (n = 1)	0	1	0	0	0	0	1	
	Negative subscale (n = 3)	1	2	0	0	0	0	3	
	General psychopathology subscale (n = 1)	0	1	0	0	0	0	1	
Internal consistency	PANSS total score (n = 13)	0	0	0	13	0	0	13	
	Positive subscale (n = 15)	14	0	1	0	0	0	15	
	Negative subscale (n = 16)	15	0	1	0	0	0	16	
	General psychopathology subscale (n = 11)	10	0	1	0	0	0	11	
Cross-cultural validity	PANSS total score (n = 4)	0	0	3	1	0	0	4	
	Positive subscale (n = 6)	0	0	3	3	0	0	6	
	Negative subscale (n = 6)	0	0	3	3	0	0	6	
	General psychopathology subscale (n = 6)	0	0	3	3	0	0	6	
Interrater reliability	PANSS total score (n = 13)	4	6	3	0	13	0	0	
	Positive subscale (n = 18)	4	4	10	0	18	0	0	
	Negative subscale (n = 18)	4	4	10	0	18	0	0	
	General psychopathology subscale (n = 17)	4	3	10	0	17	0	0	
Test-retest reliability	PANSS total score (n = 3)	0	0	2	1	3	0	0	
	Positive subscale (n = 2)	0	0	2	0	2	0	0	
	Negative subscale (n = 2)	0	0	2	0	2	0	0	
	General psychopathology subscale (n = 2)	0	0	2	0	2	0	0	
Measurement error	PANSS total score (n = 4)	0	0	1	3	0	0	4	
	Positive subscale (n = 4)	0	0	1	3	0	0	4	
	Negative subscale (n = 4)	0	0	1	3	0	0	4	
	General psychopathology subscale (n = 4)	0	0	1	3	0	0	4	
Criterion validity	PANSS total score (n = 4)	4	0	0	0	4	0	0	
Hypothesis testing for construct validity	PANSS total score (n = 7)	1	1	5	0	5	2	0	
	Positive subscale (n = 38)	7	24	6	1	36	1	1	
	Negative subscale (n = 40)	7	25	6	2	37	2	1	
	General psychopathology subscale (n = 20)	1	15	2	2	16	4	0	
Responsiveness	PANSS total score (n = 6)	3	1	2	0	6	0	0	

^aClinROM development rating resulted from *Criteria for content validity*; 2 of the 95 (-) exploratory-derived-model-ratings result from Generalized Structured Component Analysis (GSCA) and exploratory structured equation modeling (ESEM), which are not considered in the COSMIN framework. As those studies (models) fail to replicate a three-factor structure (like the original PANSS model), and EFA models are rated (-) as soon as 1 criterion is not fulfilled, those more specific structural equation modeling (SEM) studies are still rated 'insufficient'. ^bSufficient. ^cInsufficient. ^dIndeterminate. ^eInconsistent.

Table 2: COSMIN risk of bias and criteria results.

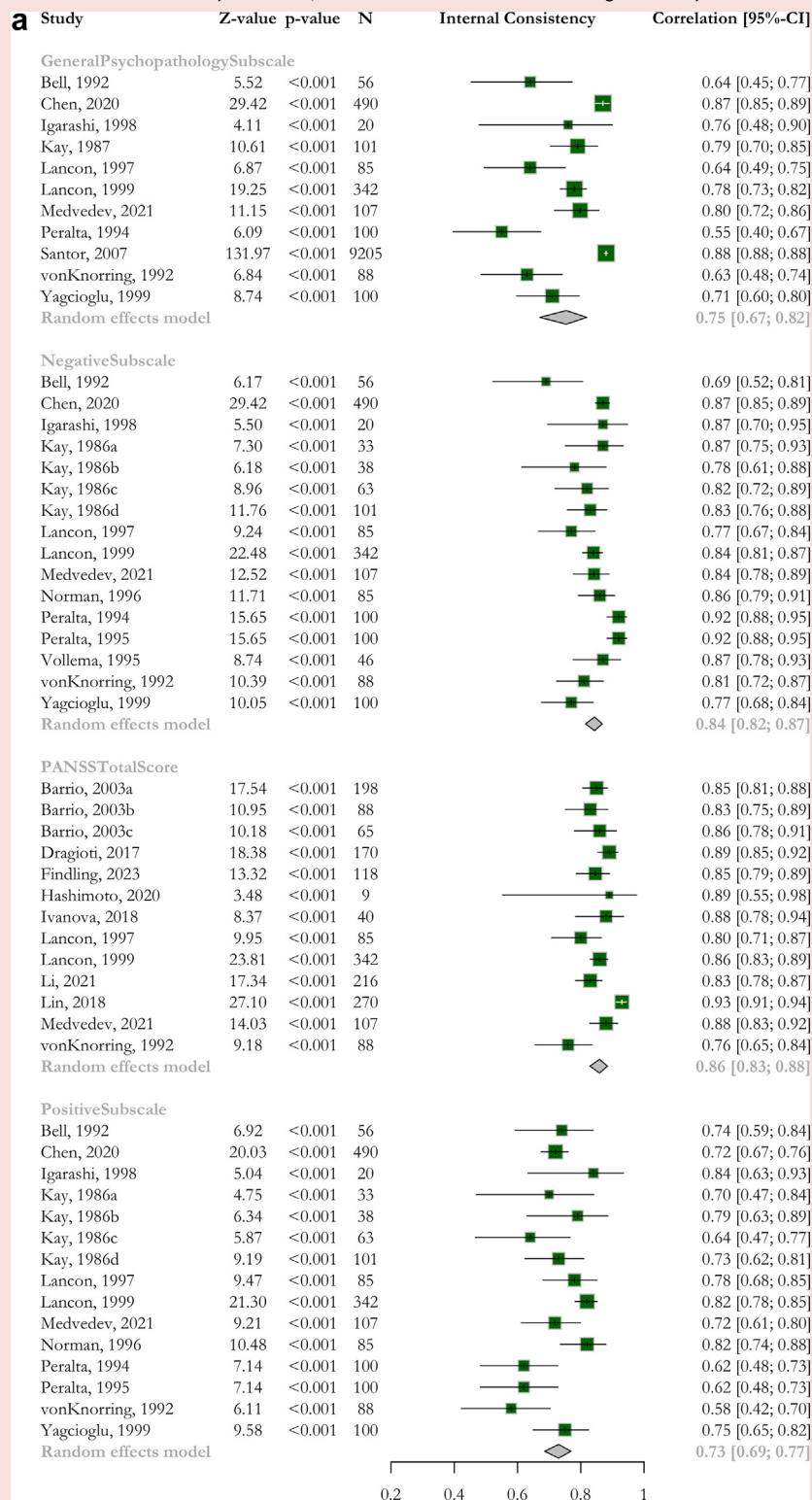
structure were identified.^{5,12,23–26,30–47,50–54,56–70,71–90,122,128–134} Their assessment could not confirm this original structure resulting in 111 “insufficient” ratings. Confirmatory-fit-indices (comparative fit indices) ranged between 0.517 and 0.920. 22 studies applied Item-Response-Theory/Rasch analyses to the original three subscale structure set up by Kay et al. (1987) which did not fulfill all necessary *Updated criteria* for a (+)/(-) rating and were therefore rated as “indeterminate”.^{1,11,21,22,27–29,38,48,49,55,135} Consequently, the overall rating on the structural validity of the original three subscale structure is “insufficient”. 17 of the 22 studies assessed the scalability of the PANSS. 15 studies conclude that the PANSS is not scalable, thus supporting the

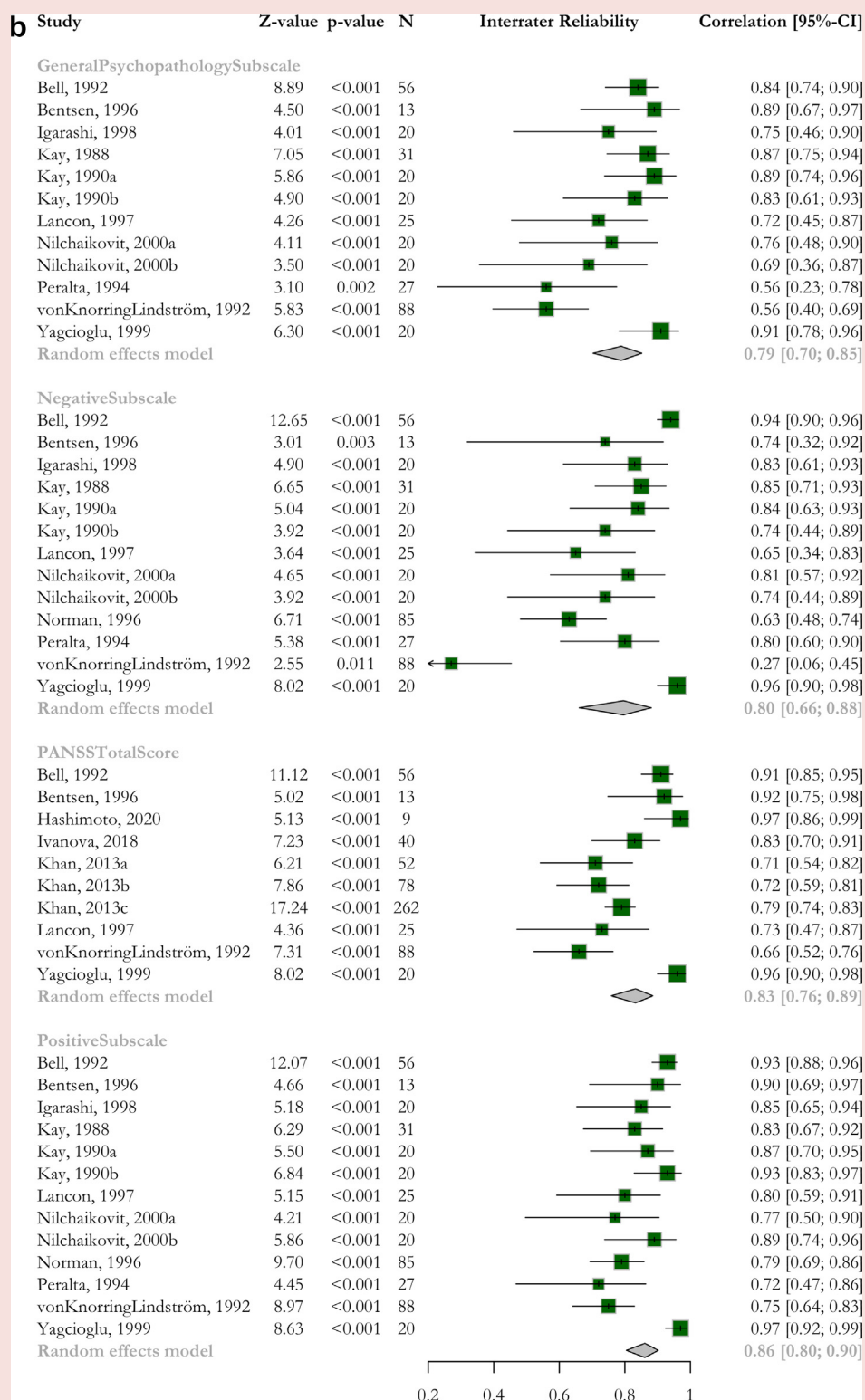
evidence for insufficiency of the structural validity of the PANSS.^{11,22,27–29} The lack of scalability of the PANSS, as demonstrated by these 15 studies, essentially refers to its failure to fit the Rasch model.

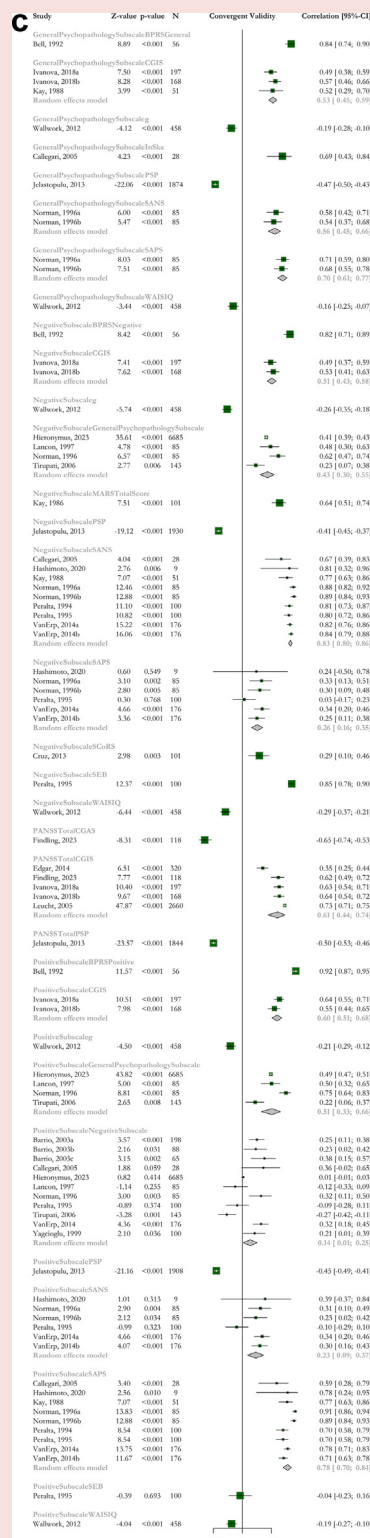
Four PCAs assessing the dimensionality of the subscales and one Rasch analysis of the negative subscales' scalability were identified.^{49,91,92} All five assessments on structural validity of the subscales were rated as “indeterminate” because the assessment methods used are not considered by COSMIN (COSMIN only provides unidimensionality criteria for IRT/Rasch analyses) or because not all criteria required by COSMIN were present (for IRT/Rasch analyses, COSMIN requires 4

Panel 1: Results from meta analysis.

Standard random-effects meta-analyses were conducted using Comprehensive Meta-Analysis Version 2; Panel 1a) Cronbach's alphas for *internal consistency*, Panel 1b) correlation coefficients for *interrater reliability*, Panel 1c) correlation coefficients for *convergent validity*.







criteria to be met: no violation of unidimensionality, of local independence and of monotonicity and adequate model fit for a “sufficient” rating. Only one of them was met). The assessments delivered inconsistent results on the unidimensionality of the negative subscale.^{49,92} A lack of scalability of the negative subscale and of unidimensionality for the general psychopathology scale were reported.^{49,91} The unidimensionality assessment of the positive subscale concluded that it is unidimensional.⁴⁹

Internal consistency. 55 studies on internal consistency were identified.^{1,21,29,37,55,74,80,82,93–108} At least low evidence for sufficient structural validity is a COSMIN requirement for a “sufficient” rating on internal consistency.^{8–10} Thus, even though the pooled Cronbach’s alpha values of $\alpha = 0.730$ (positive subscale), $\alpha = 0.844$ (negative subscale), $\alpha = 0.754$ (general psychopathology subscale) and $\alpha = 0.859$ (PANSS total) surpass the threshold of $\alpha \geq 0.70$ which is deemed as acceptable by COSMIN and satisfies the common notion of using the threshold of $\alpha \geq 0.70$ for an instrument to have an acceptable level of internal consistency, the internal consistency ratings of the original three subscale structure of the PANSS and its subscales were “indeterminate”.^{8–10,136,137}

Remaining measurement properties. For the remaining measurement properties (cross-cultural validity, reliability, measurement error, criterion validity, construct validity, and responsiveness) the ratings apply for both the PANSS total and all three subscales, except for criterion validity and responsiveness, for which only evidence on the PANSS total was found.

Cross-cultural validity/measurement invariance. 22 studies on cross-cultural validity could be identified, while none were available on measurement invariance.^{109–112} Percentage differences of least square means from Analysis of Covariance and Pearson correlation coefficients were used to compare the PANSS in American versus Chinese, as well as between Spanish and English PANSS assessments. Percentage differences of least square means (LSM) ranged between 0.1% and 5.1% in one publication and between 8.8% and 37.5% in a second.^{111,112} Aggarwal et al. ascribe the differences between the Chinese and American samples mainly to cultural differences.^{111,112} Pooled Pearson correlations of the subscales were strong and ranged between 0.77 and 0.905, indicating satisfactory cross-cultural validity.^{109,110,138} Nonetheless, COSMIN requires other tests to assess cross-cultural validity than simple Pearson correlations and percentage differences of LSM. Thus, cross-cultural validity is rated as “indeterminate”.^{8–10}

Measurement error. The identified 16 studies on measurement error used G-theory and provided moderate to high generalizability-coefficients between 0.647 and 0.93, indicating that a moderate to large proportion of the variance in the observed scores is due to the true

score variance, not due to error variance.^{98,116,139,140} However, G-theory is not a method considered by COSMIN to assess measurement error, which was therefore rated as “indeterminate” as well.^{8–10}

Reliability and criterion validity. The pooled results for studies on interrater reliability ($n_{\text{studies}} = 66$; $r_{\text{pooled-range}} = 0.790\text{--}0.885$), test-retest reliability ($n_{\text{studies}} = 9$; $r_{\text{pooled-range}} = 0.737\text{--}0.797$), and criterion validity ($n_{\text{studies}} = 4$; $r_{\text{pooled}} = 0.877$), surpass the *Updated criteria* criterion of $r \geq 0.7$, resulting in reliability and criterion validity ratings of “sufficient”.^{1,8–10,80,82,93,95,96,98,101,102,105–107,109,110,113–118}

Construct validity. Known groups validity: Two studies assessing the ability of the scales to discriminate between different disease stages were identified.⁹⁹ Analysis of Variance was used, which is not considered by COSMIN, resulting in “indeterminate” ratings, even though the authors report a lack of group differences on control variables, indicating good discriminant validity.^{8–10,99}

Convergent and divergent validity: 105 studies correlating PANSS scores to measures of schizophrenia symptom severity, neurocognition, functioning, global assessment, and other more specific measurement scales were identified.^{21,60,80,82,92,94–96,99–101,103,105,107,117–124}

A sufficient percentage of hypotheses (71.43%–97.30%) was confirmed to rate construct validity as “sufficient”.

Responsiveness. As all hypotheses for responsiveness ($n_{\text{studies}} = 6$) were confirmed, responsiveness was rated as “sufficient”.^{29,120,125–127}

Publication bias. Internal consistency, interrater reliability, and the correlation between the positive and negative subscale were investigated for small-trial/publication bias, and no indications for such bias were identified.

GRADE approach

For the PANSS total, the qualities of the evidence on the structural validity, reliability, criterion validity, and responsiveness was graded as “high”, on the cross-cultural validity and construct validity as “moderate” and on the content validity, internal consistency and measurement error as “low”.

The quality of the evidence for the subscales was graded as “high” regarding internal consistency, reliability and construct validity, as “moderate” to “high” regarding structural validity, as “moderate” regarding cross-cultural validity, and as “low” regarding measurement error.

The detailed results on the quality of evidence together with the reasons for downgrading are presented in [Table 3](#).

Recommendations

The PANSS total is categorized as “not-recommendable” (C), the subscales fall in the “potential to be recommended” (B) category.

The PANSS total score is rated lower than the subscales due to strong evidence for “insufficient” structural validity. The subscales are not necessarily better, but there is less evidence available, which places them in the intermediate category “B”, as future evidence could hypothetically lead to a more positive evaluation. This is due to COSMIN considering each subscale as a scale of its own, requiring individual assessments. However, the categorization of the subscales in the intermediate category “B” needs to be interpreted in the context of the lacking structural validity of the PANSS total score, which does not suggest a three-factor structure.

Feasibility

Strengths of the PANSS are its availability in over 40 languages, simple score calculation by summing up item scores, the availability of the PANSS manual with its precise definitions, the Structured Clinical Interview for PANSS assessments (SCI-PANSS) and the Informant Questionnaire (IQ-PANSS).^{141–143} Weaknesses are the scale’s long completion time of 30–50 min, requirements for rater training to achieve acceptable interrater reliability making the standardization process laborious, and the high cost of the instrument (\$96,40 for the PANSS Technical Manual, \$90,80 for 25 PANSS Rating and Profile Forms) due to copyright reasons.^{1,114,144} The European Medicines Agency (EMA) requires the use of the PANSS for drug approval (compare [Appendix S8](#)).¹⁴⁵

Discussion

This systematic COSMIN review and meta-analysis based on 118 publications showed that the PANSS is a strong scale in terms of reliability, construct validity, criterion validity, and responsiveness, which are important for treatment trials. Nevertheless, it also has major shortcomings. Its content validity is questionable because for the development of the PANSS, the BPRS was simply supplemented by items of the PRS rather than using a methodological scale construction process (e.g., derive the items statistically from a large item pool, involve clinicians and patients in the process, etc.). Structural validity is poor because the original 3 factor model has a bad statistical fit.

The satisfying responsiveness currently is a justification for the use of the PANSS in clinical trials and for the European Medicines Agency (EMA) as a requirement for drug approval.¹⁴⁵ The ability to separate between drug and placebo (an aspect of responsiveness) is of key importance in such trials. Nevertheless, even very short scales such as the Clinical Global Impression Scale regularly separate between interventions.¹⁴⁶ Other advantages of the PANSS are that its scores are relatively interpretable, although the linking analyses which made this possible were carried out late after its introduction.^{119,125,126,147–150}

	Summary or pooled result	Overall rating	Quality of evidence
Content validity			
PANSS-30 total score	Relevance rating: ±, Comprehensiveness rating: -, Comprehensibility rating: ± (overall ratings resulting from the PROM development study and the reviewer rating)	Inconsistent	Low (no content validity studies available)
Structural validity			
Positive subscale	Unidimensional score	Indeterminate	Moderate (just one adequate study available)
Negative subscale	1 out of 2 PCA analyses result in unidimensionality; 1/1 scalability assessments reject scalability	Indeterminate	High
General psychopathology subscale	No unidimensional score	Indeterminate	Moderate (one adequate study available)
PANSS-30 total score	None of the EFA derived models fulfill the criteria; CFI-range: 0.517–0.920 (total sample size: 6114), RMSEA-range: 0.108–0.123 (total sample size: 2644), 15/17 scalability assessments reject scalability	Insufficient	High
Internal consistency			
Positive subscale	Pooled: Cronbach's α = 0.730 (total sample size: 1808)	Indeterminate (15× (?))	High
Negative subscale	Pooled: Cronbach's α = 0.844 (total sample size: 1854)	Indeterminate (16× (?))	High
General psychopathology subscale	Pooled: Cronbach's α = 0.754 (total sample size: 10,694)	Indeterminate (11× (?))	High
PANSS-30 total score	Pooled: Cronbach's α = 0.859 (total sample size: 1796)	Indeterminate (13× (?))	Low (only inadequate studies available)
Cross-cultural validity			
Positive subscale	Pooled: r = 0.905 (total sample size: 37); percentage difference range: 0.6–37.5%	Indeterminate (6× (?))	Moderate (multiple doubtful studies available)
Negative subscale	Pooled: r = 0.776 (total sample size: 37); percentage difference range: 0.1%–8.8%	Indeterminate (6× (?))	Moderate (multiple doubtful studies available)
General psychopathology subscale	Pooled: r = 0.770 (total sample size: 37); percentage difference range: 5.1–34.3%	Indeterminate (6× (?))	Moderate (multiple doubtful studies available)
PANSS-30 total score	Percentage difference range: 5.1–26.1%	Indeterminate (4× (?))	Moderate (multiple doubtful studies available)
Reliability			
Positive subscale	Pooled: interrater reliability: r = 0.885 (total sample size: 596), test-retest reliability: r = 0.791 (total sample size: 122)	Sufficient	High
Negative subscale	Pooled: interrater reliability: r = 0.790 (total sample size: 599), test-retest reliability: r = 0.762 (total sample size: 122)	Sufficient	High
General psychopathology subscale	Pooled: interrater reliability: r = 0.828 (total sample size: 514), test-retest reliability: r = 0.737 (total sample size: 122)	Sufficient	High
PANSS-30 total score	Pooled: interrater reliability: r = 0.812 (total sample size: 701), test-retest reliability: r = 0.797 (total sample size: 1771)	Sufficient	High
Measurement error			
Positive subscale	Generalizability coefficient range: 0.662–0.80 (total sample size: 319)	Indeterminate (4× (?))	Low (multiple inadequate studies available)
Negative subscale	Generalizability coefficient range: 0.647–0.77 (total sample size: 319)	Indeterminate (4× (?))	Low (multiple inadequate studies available)
General psychopathology subscale	Generalizability coefficient range: 0.691–0.84 (total sample size: 319)	Indeterminate (4× (?))	Low (multiple inadequate studies available)
PANSS-30 total score	Generalizability coefficient range: 0.687–0.93 (total sample size: 319)	Indeterminate (4× (?))	Low (multiple inadequate studies available)
Criterion validity			
PANSS-30 total score	Pooled: r = 0.877 (total sample size = 3415)	Sufficient	High
Hypothesis testing			
Positive subscale	36 out of 37 hypotheses confirmed	Sufficient	High
Negative subscale	37 out of 39 hypotheses confirmed	Sufficient	High
General psychopathology subscale	16 out of 20 hypotheses confirmed	Sufficient	High
PANSS-30 total score	5 out of 7 hypotheses confirmed	Sufficient	Moderate (due to inconsistent results)
Responsiveness			
PANSS-30 total score	6 out of 6 hypotheses confirmed	Sufficient	High

Table 3: COSMIN Summary of findings.

Such analyses have shown that approximately 25% PANSS total score reduction from baseline means minimally improved according to the CGI, 50% much

improved, and 15/5 absolute PANSS total/negative points reduction minimally improved according to the CGI.^{119,125,148–150} It should be noted that the 30 minimum

points of the PANSS need to be subtracted when calculating percentage improvement because otherwise response is underestimated.^{150–152} Other strong points are the availability in more than 40 languages, a PANSS institute which among others provides training, and interview guides such as the SCI-PANSS, a structured clinical interview, and the IQ-PANSS, a structured informant questionnaire.^{141–143,153}

One of the most frequently assessed and questioned aspects of the PANSS is its structural validity. Exploratory approaches extensively applied principal component analyses and exploratory factor analyses, but also more robust and more valid methods such as 10-fold-cross-validations, which clearly demonstrated the inaccuracy of the three-subscale structure.^{42,59} These analyses yielded a variety of models with factor numbers ranging between 4 and 9, but most analyses supported a five-factor-model solution. We will report our assessment of these factor models in a separate publication. We emphasize that the usefulness of the original negative subscale is particularly low. In the NIMH-MATRICES consensus, it was decided that negative symptoms encompass 5 domains (Blunted affect, Alogia, Asociality, Avolition, Anhedonia), of which the PANSS covers only the first three.¹⁵⁴ Moreover, studies have shown that the negative subscale is not scalable and not unidimensional.^{91,92} In recent reviews, the modern scales for negative symptoms, Clinical Assessment Interview for Negative Symptoms (CAINS) and Brief negative Symptom Scale (BNSS), were overall better than older scales as the Scale for the Assessment of Negative Symptoms (SANS) and the Negative Symptom Assessment (NSA).^{13,14,155–160} The PANSS negative subscale is also inferior to the CAINS and BNSS for the reasons just mentioned.

Due to unclear or inadequate structural validity, internal consistency is questionable as well. According to COSMIN good structural validity is a prerequisite for the interpretability of internal consistency statistics.^{8–10} High Cronbach's alpha values indicate a high interrelatedness of the items of the PANSS and its subscales, but if the structure of a scale is unclear the informative value is limited because the underlying structure may in reality be different. Good internal consistency is a prerequisite for a good ("A") recommendation by COSMIN.^{8–10} Thus, if, for example, improved statistical models found a factor structure of the PANSS with better fit, the recommendability of the scale could increase.

These weaknesses have also implications on the interpretation of previous studies. In particular, as long as the content validity of a scale is not clear, we cannot be sure that we are asking the right questions. Having said this, the PANSS has not been designed as a diagnostic scale. Its main purpose is to be sensitive for change ("responsiveness") and this is a strength of the PANSS. Nevertheless, PANSS total score-based effect sizes of antipsychotic drugs compared to placebo are

relatively low (average ~0.5; compare Leucht et al. (2017)) and might be higher with a better scale. In terms of structural validity, some studies have analyzed both the original 3 subscale structure and a 5-factor structure but did not find clear differences between them (e.g., Nemeth et al. (2017)).^{161,162} The PANSS does not include all domains found important by the NIMH consensus for negative symptoms. Thus, previous studies could not assess all these domains. In sum if a new scale were developed, it needs to be found out whether it would reach better results.

Finally, feasibility is problematic. An interview lasts about 45 min which precludes the use of the PANSS in routine clinical practice.¹ The anchors for the ratings are clearly defined but as there are 30 items (i.e., 30 times 6 degrees = 180 definitions), only very experienced raters know them by heart.¹⁴² Moreover, it has been shown that not all items are scalable, i.e., not all items contribute additional information.^{11,22,27–29} This has led to the development of various short versions, such as the 14-item Mini-PANSS, the adapted 19-item Mini-PANSS, the PANSS-8, the six-item Brief PANSS, the PANSS-6,^{11,29,49,55,163–165} and the Brief Evaluation of Psychosis Symptom Domains (BE-PSD) scale.¹⁶⁶ Five-factor solutions and PANSS-6 will be addressed in our companion reviews.¹⁶⁷

This systematic review has several limitations.

First, while the COSMIN framework offers a structured approach to evaluating the measurement properties of instruments like the PANSS, it does not accommodate certain methodologies that might provide a more precise analysis. This required us to use certain workarounds to accommodate for the variety of approaches used and prevented us from assessing two studies that applied network analyses and orthonormal projective non-negative matrix factorization (OPNMF), both modern ways to assess scale structure.^{108,168} The first study found that the interconnection of symptoms appears to be different between treatment responsive versus treatment resistant patients.¹⁶⁸ The second one supported a four-factor structure (see [Appendix S9](#)).¹⁰⁸ Moreover, some old methods which have been frequently used are increasingly considered less appropriate (e.g., Cronbach's alpha for reliability).¹⁶⁹

Thus, several methodologically different procedures with different levels of validity were assessed based on the same criteria, e.g., no consideration of different rotation methods or eigenvalue criteria, or different kinds of correlation coefficients entered calculations despite their differences.¹⁷⁰ Nevertheless, such studies which could not be implemented in the COSMIN framework and were, therefore, rated as indeterminate ("?"), yielded similar results. Only very few analyses from the 119 included studies could not be incorporated into the COSMIN framework, therefore their impact is minor. We recommend that in future versions of COSMIN new methods to assess rating scales will be implemented.

Another point is that the COSMIN framework mainly bases its final step “recommendation” on the domains “content validity” and “internal consistency”. Some may set this focus differently.

Second, the PANSS is an instrument with ordinal scaling, but as the original authors of the included studies we assumed that it is continuous. This problem applies to many psychiatric rating scales, but the implications are unclear.

Third, the populations, raters (degree of training) and the methods used in the studies were quite heterogeneous.

Fourth, we report a lack of evidence on responsiveness of the subscales in our review as none of the included validation studies assessed it. However, multiple regular studies applying the PANSS subscales could demonstrate its ability to detect pre- and post-treatment differences, but these would not have changed the overall rating of the PANSS.^{162,171}

Fifth, our choice to consider the BPRS as gold standard for criterion validity might be questioned, as the PANSS was derived from the BPRS. If we had not considered the BPRS as the gold standard, the domain criterion validity would not have been assessable by COSMIN. Such a situation does not lead to a downgrading. Thus, the overall judgment would not have changed.

Sixth, although COSMIN is a highly elaborated and operationalized framework to systematically review the measurement properties of rating scales, there is a degree of subjectivity.^{8–10}

The PANSS is a comprehensively evaluated tool with strengths in responsiveness, reliability and construct validity. The field also has vast experience with this scale. Nevertheless, there are important shortcomings primarily regarding content- and structural validity, which are the fundamental domains in the COSMIN approach. If only the strengths in terms of responsiveness and reliability of the PANSS are needed, it can be useful, but its weaknesses limit interpretation because content and structural validity determine the meaning of a score. Moreover, the length of the scale limits its use in daily clinical routine and in clinical trials. The original three-subscale structure is not valid. In addition to the short versions of the PANSS mentioned above, short scales such as the Brief Evaluation of Psychosis Symptom Domains (BE-PSD) and the Clinical Global Impression-Schizophrenia Scale may be more applicable in practice.^{166,172,173} Appropriately developed new scales for comprehensively and accurately assessing the symptoms of schizophrenia are warranted, possibly within a formative rather than reflective framework.

Contributors

S.G. and M.R. contributed equally to this work and are both recognized as first authors. S.G. and M.R. had full access to all data in the study and take responsibility for verification of the data, and the accuracy of the data analysis. The authors confirm contribution to the paper as follows: study conception and design: S.G., M.R., and S.L.; data collection: S.G. and M.R.; analysis and interpretation of results: S.G., M.R., and S.L.;

statistical consultation: M.B.; draft manuscript preparation: S.G. and M.R.; revision for important intellectual content: S.L., M.B., S.W., L.W., J.P., and J.M.D.; The work will be part of the doctoral thesis of S.G. and M.R.; All authors reviewed the results and approved the final version of the manuscript. All authors have agreed to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, are appropriately investigated, resolved, and the resolution documented in the literature.

Data sharing statement

Underlying study data will be shared by the corresponding author (stefan.leucht@tum.de) on reasonable request for academic and research purposes and is subject to data sharing agreements. Supplementary data can be accessed in the [Appendix](#).

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT-4o by OpenAI, Inc. to improve readability and comprehensibility. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of interests

In the last 3 years, Dr Leucht has received honoraria for advising, consulting, and for lectures and/or for educational material from Angelini, Apsen, Boehringer Ingelheim, Eisai, Ekademia, Gedeon-Richter, Janssen, Karuna, Kynexis, Lundbeck, Medichem, Medscape, Mitsubishi, Neurotorium, Otsuka, Novo Nordisk, Recordati, Rovi, and Teva. All other authors declare that they have no conflicts of interest relevant to this research.

Acknowledgements

We are grateful to the authors who replied to our mails/inquiries: Jimmy Lee Chee Keong, Jan Stochl, Leslie Citrome, Danilo Bzdok, Dalibor Karlović, Ji Chen, Helio Elkis, Eric Youngstrom, Stephen Marder, Jean-Pierre Lindenmayer, Daniel Purnine, Morris D. Bell, Kenneth Koblan, Wieneke Mokkink, Alessandro Rossi, Kim Good.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.eclinm.2025.103155>.

References

- Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull.* 1987;13(2):261–276.
- Liechti S, Capodilupo G, Opler DJ, Opler M, Yang LH. A developmental history of the positive and negative syndrome scale (PANSS). *Innov Clin Neurosci.* 2017;14(11–12):12–17.
- Crow TJ. Molecular pathology of schizophrenia: more than one disease process? *Br Med J.* 1980;280(6207):66–68.
- Crow TJ. Positive and negative schizophrenic symptoms and the role of dopamine. *Br J Psychiatry.* 1980;137:383–386.
- Mass R, Schoemig T, Hitschfeld K, Wall E, Haasen C. Psychopathological syndromes of schizophrenia: evaluation of the dimensional structure of the positive and negative syndrome scale. *Schizophr Bull.* 2000;26(1):167–177.
- Overall JE, Gorham DR. The brief psychiatric rating scale. *Psychol Rep.* 1962;10:799–812.
- Singh MM, Kay SR. A comparative study of haloperidol and chlorpromazine in terms of clinical effects and therapeutic reversal with benztropine in schizophrenia. Theoretical implications for potency differences among neuroleptics. *Psychopharmacologia.* 1975;43(2):103–113.
- Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018;27(5):1171–1179.
- Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018;27(5):1147–1157.

- 10 Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res.* 2018;27(5):1159–1170.
- 11 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(6):436–444.
- 12 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. *J Clin Psychiatry.* 1997;58(12):538–546.
- 13 Weigel L, Wehr S, Galderisi S, et al. The Brief negative Symptom Scale (BNSS): a systematic review of measurement properties. *Schizophrenia (Heidelb).* 2023;9(1):45.
- 14 Wehr S, Weigel L, Davis J, Galderisi S, Mucci A, Leucht S. Clinical assessment interview for negative symptoms (CAINS): a systematic review of measurement properties. *Schizophr Bull.* 2024;50(4):747–756.
- 15 Roithmeier M, Geck S, Peter N, Leucht S. Systematic review of the positive and negative syndrome scale following the COSMIN standard. <https://osf.io/2t593/>; 2024. Accessed August 26, 2024.
- 16 Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res.* 2009;18(8):1115–1123.
- 17 Hedlund JL. The brief psychiatric rating scale (BPRS): a comprehensive review. *J Oper Psychiatr.* 1980;11:48–65.
- 18 Elsman EBM, Mokkink LB, Langendoen-Gort M, et al. Systematic review on the measurement properties of diabetes-specific patient-reported outcome measures (PROMs) for measuring physical functioning in people with type 2 diabetes. *BMJ Open Diabetes Res Care.* 2022;10(3):e002729.
- 19 Zúñiga Le-Bert M, Wiessner M, Wehr S, Weigel L, Leucht S. Schizophrenia quality of life scale and schizophrenia quality of life scale revision 4: a systematic review of measurement properties. *Schizophr Bull.* 2024. <https://doi.org/10.1093/schbul/sbae119>.
- 20 Borenstein M, Hedges L, Higgins J, Rothstein H. *An Introduction to Meta-analysis.* West Sussex, United Kingdom: John Wiley & Sons Ltd; 2009;19.
- 21 Findling RL, Youngstrom EA, McClellan JM, et al. An optimized version of the positive and negative symptoms scale (PANSS) for pediatric trials. *J Am Acad Child Adolesc Psychiatry.* 2023;62(4):427–434.
- 22 Baandrup L, Allerup P, Nielsen MO, et al. Scalability of the Positive and Negative Syndrome Scale in first-episode schizophrenia assessed by Rasch models. *Acta Psychiatr Scand.* 2022;146(1):21–35.
- 23 Lim K, Peh OH, Yang Z, et al. Large-scale evaluation of the positive and negative syndrome scale (PANSS) symptom architecture in schizophrenia. *Asian J Psychiatr.* 2021;62:102732.
- 24 Fountoulakis KN, Dragioti E, Theofilidis AT, et al. Staging of schizophrenia with the use of PANSS: an international multi-center study. *Int J Neuropsychopharmacol.* 2019;22(11):681–697.
- 25 Walsh-Messinger J, Antonius D, Opler M, et al. Factor structure of the positive and negative syndrome scale (PANSS) differs by sex. *Clin Schizophr Relat Psychoses.* 2018;11(4):207–213.
- 26 Petruzzelli MG, Margari L, Bosco A, Craig F, Palumbi R, Margari F. Early onset first episode psychosis: dimensional structure of symptoms, clinical subtypes and related neurodevelopmental markers. *Eur Child Adolesc Psychiatry.* 2018;27(2):171–179.
- 27 Ostergaard SD, Foldager L, Mors O, Bech P, Correll CU. The validity and sensitivity of PANSS-6 in treatment-resistant schizophrenia. *Acta Psychiatr Scand.* 2018;138(5):420–431.
- 28 Ostergaard SD, Foldager L, Mors O, Bech P, Correll CU. The validity and sensitivity of PANSS-6 in the clinical antipsychotic trials of intervention effectiveness (CATIE) study. *Schizophr Bull.* 2018;44(2):453–462.
- 29 Lin CH, Lin HS, Lin SC, Kuo CC, Wang FC, Huang YH. Early improvement in PANSS-30, PANSS-8, and PANSS-6 scores predicts ultimate response and remission during acute treatment of schizophrenia. *Acta Psychiatr Scand.* 2018;137(2):98–108.
- 30 Lefort-Besnard J, Varoquaux G, Derntl B, et al. Patterns of schizophrenia symptoms: hidden structure in the PANSS questionnaire. *Transl Psychiatry.* 2018;8(1):237.
- 31 Hopkins SC, Ogirala A, Loebel A, Koblan KS. Transformed PANSS factors intended to reduce pseudospecificity among symptom domains and enhance understanding of symptom change in antipsychotic-treated patients with schizophrenia. *Schizophr Bull.* 2018;44(3):593–602.
- 32 Grover S, Dua D, Chakrabarti S, Avasthi A. Factor analysis of symptom dimensions (psychotic, affective and obsessive compulsive symptoms) in schizophrenia. *Asian J Psychiatr.* 2018;38:72–77.
- 33 De Freitas R, Dos Santos B, Altamura C, et al. Can the positive and negative syndrome scale (PANSS) differentiate refractory from non-refractory schizophrenia? A factor analytic investigation based on data from the pattern cohort study. *Schizophr Bull.* 2018;44 (Supplement 1):S371.
- 34 Anderson AE, Marder S, Reise SP, et al. Bifactor modeling of the positive and negative syndrome scale: generalized psychosis spans schizoaffective, bipolar, and schizophrenia diagnoses. *Schizophr Bull.* 2018;44(6):1204–1216.
- 35 Yehya A, Ghuloum S, Mahfoud Z, et al. Validation of the five-factor model of the Arabic version of the positive and negative syndrome scale in schizophrenia. *Psychopathology.* 2017;50(3):211–218.
- 36 Peitl V, Štefanović M, Karlovič D. Depressive symptoms in schizophrenia and dopamine and serotonin gene polymorphisms. *Prog Neuropsychopharmacol Biol Psychiatry.* 2017;77:209–215.
- 37 Dragioti E, Wiklund T, Siamouli M, Moutou K, Fountoulakis KN. Could PANSS be a useful tool in the determining of the stages of schizophrenia? A clinically operational approach. *J Psychiatr Res.* 2017;86:66–72.
- 38 Anderson AE, Reise SP, Marder SR, Mansolf M, Han C, Bilder RM. Disparity between general symptom relief and remission criteria in the positive and negative syndrome scale (PANSS): a post-treatment bifactor item response theory model. *Innov Clin Neurosci.* 2017;14(11–12):41–53.
- 39 Thokagevistk K, Millier A, Lenert L, Sadikhov S, Moreno S, Toumi M. Validation of disease states in schizophrenia: comparison of cluster analysis between US and European populations. *J Market Access Health Policy.* 2016;4. <https://doi.org/10.3402/jmahp.v4.30725>.
- 40 Best MW, Grossman M, Oyewumi LK, Bowie CR. Examination of the Positive and Negative Syndrome Scale factor structure and longitudinal relationships with functioning in early psychosis. *Early Interv Psychiatry.* 2016;10(2):165–170.
- 41 Xu K, Krystal JH, Ning Y, et al. Preliminary analysis of positive and negative syndrome scale in ketamine-associated psychosis in comparison with schizophrenia. *J Psychiatr Res.* 2015;61:64–72.
- 42 Anderson A, Wilcox M, Savitz A, et al. Sparse factors for the positive and negative syndrome scale: which symptoms and stage of illness? *Psychiatry Res.* 2015;225(3):283–290.
- 43 Stochl J, Jones PB, Plaistow J, et al. Multilevel ordinal factor analysis of the positive and negative syndrome scale (PANSS). *Int J Methods Psychiatr Res.* 2014;23(1):25–35.
- 44 Higuchi CH, Ortiz B, Berberian AA, et al. Factor structure of the positive and negative syndrome scale (PANSS) in Brazil: convergent validation of the Brazilian version. *Br J Psychiatry.* 2014;36(4):336–339.
- 45 Jiang J, Sim K, Lee J. Validated five-factor model of positive and negative syndrome scale for schizophrenia in Chinese population. *Schizophr Res.* 2013;143(1):38–43.
- 46 Kumar A, Khess CRJ. Factor analysis of positive and negative syndrome scale in schizophrenia: an exploratory study. *Indian J Psychiatry.* 2012;54(3):233–238.
- 47 Kim JH, Kim SY, Lee J, Oh KJ, Kim YB, Cho ZH. Evaluation of the factor structure of symptoms in patients with schizophrenia. *Psychiatry Res.* 2012;197(3):285–289.
- 48 Levine SZ, Rabinowitz J, Rizopoulos D. Recommendations to improve the positive and negative syndrome scale (PANSS) based on item response theory. *Psychiatry Res.* 2011;188(3):446–452.
- 49 Khan A, Lewis C, Lindenmayer JP. Use of non-parametric item response theory to develop a shortened version of the Positive and Negative Syndrome Scale (PANSS). *BMC Psychiatry.* 2011;11:178.
- 50 Citrome L, Meng X, Hochfeld M. Efficacy of iloperidone in schizophrenia: a PANSS five-factor analysis. *Schizophr Res.* 2011;131(1–3):75–81.
- 51 Hwang SSH, Chang JS, Lee KY, Ahn YM, Kim YS. The causal model of insight in schizophrenia based on the positive and negative syndrome scale factors and the structural equation modeling. *J Nerv Ment Dis.* 2009;197(2):79–84.
- 52 Gil D, Bengochea R, Arrieta M, et al. Validity of the cognitive factor of the Positive and Negative Syndrome Scale as a measure of cognitive functioning in schizophrenia. *Rev Psiquiatr Salud Ment.* 2009;2(4):160–168.
- 53 Ruiz-Veguilla M, Cervilla JA, Barrigon ML, et al. Neurodevelopmental markers in different psychopathological dimensions

- of first episode psychosis: the ESPIGAS study. *Eur Psychiatry*. 2008;23(8):533–540.
- 54 Lindenmayer JP, Bossie CA, Kujawa M, Zhu Y, Canuso CM. Dimensions of psychosis in patients with bipolar mania as measured by the positive and negative syndrome scale. *Psychopathology*. 2008;41(4):264–270.
- 55 Santor DA, Ascher-Svanum H, Lindenmayer JP, Obenchain RL. Item response analysis of the positive and negative syndrome scale. *BMC Psychiatry*. 2007;7:66.
- 56 Ruhrmann S, Kissling W, Lesch OM, Schmauss M, Seemann U, Philipp M. Efficacy of flupentixol and risperidone in chronic schizophrenia with predominantly negative symptoms. *Prog Neuropsychopharmacol Biol Psychiatry*. 2007;31(5):1012–1022.
- 57 Levine SZ, Rabinowitz J. Revisiting the 5 dimensions of the positive and negative syndrome scale. *J Clin Psychopharmacol*. 2007;27(5):431–436.
- 58 Villalta-Gil V, Vilaplana M, Ochoa S, et al. Four symptom dimensions in outpatients with schizophrenia. *Compr Psychiatry*. 2006;47(5):384–388.
- 59 van der Gaag M, Hoffman T, Remijsen M, et al. 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.
- 60 Tirupati SN, Padmavati R, Thara R, McCreedy RG. Psychopathology in never-treated schizophrenia. *Compr Psychiatry*. 2006;47(1):1–6.
- 61 Reichenberg A, Rieckmann N, Harvey PD. Stability in schizophrenia symptoms over time: findings from the Mount Sinai Pilgrim Psychiatric Center Longitudinal Study. *J Abnorm Psychol*. 2005;114(3):363–372.
- 62 Fresán A, De la Fuente-Sandoval C, Loyzaga C, et al. A forced five-dimensional factor analysis and concurrent validity of the Positive and Negative Syndrome Scale in Mexican schizophrenic patients. *Schizophr Res*. 2005;72(2–3):123–129.
- 63 Mohr PE, Cheng CM, Claxton K, et al. The heterogeneity of schizophrenia in disease states. *Schizophr Res*. 2004;71(1):83–95.
- 64 Masiak M, Loza B. Core factors of schizophrenia structure based on PANSS and SAPS/SANS results. Discerning and head-to-head comparison of PANSS and SAPS/SANS validity. *Psychiatr Pol*. 2004;38(5):795–808.
- 65 Lindenmayer JP, Czobor P, Volavka J, et al. Effects of atypical antipsychotics on the syndromal profile in treatment-resistant schizophrenia. *J Clin Psychiatry*. 2004;65(4):551–556.
- 66 Lee KH, Harris AW, Loughland CM, Williams LM. The five symptom dimensions and depression in schizophrenia. *Psychopathology*. 2003;36(5):226–233.
- 67 Honey GD, Sharma T, Suckling J, et al. The functional neuroanatomy of schizophrenic subsyndromes. *Psychol Med*. 2003;33(6):1007–1018.
- 68 Emsley R, Rabinowitz J, Torremán M, et al. The factor structure for the Positive and Negative Syndrome Scale (PANSS) in recent-onset psychosis. *Schizophr Res*. 2003;61(1):47–57.
- 69 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.
- 70 El Yazaji M, Battas O, Agoub M, et al. Validity of the depressive dimension extracted from principal component analysis of the PANSS in drug-free patients with schizophrenia. *Schizophr Res*. 2002;56(1–2):121–127.
- 71 Wolthaus JE, Dingemans PM, Schene AH, et al. Component structure of the Positive and Negative Syndrome Scale (PANSS) in patients with recent-onset schizophrenia and spectrum disorders. *Psychopharmacology (Berl)*. 2000;150(4):399–403.
- 72 Lykouras L, Oulis P, Psarros K, et al. Five-factor model of schizophrenic psychopathology: how valid is it? *Eur Arch Psychiatry Clin Neurosci*. 2000;250(2):93–100.
- 73 Lançon C, Auquier P, Nayt G, Reine G. Stability of the five-factor structure of the positive and negative syndrome scale (PANSS). *Schizophr Res*. 2000;42(3):231–239.
- 74 Lançon C, Reine G, Llorca PM, Auquier P. Validity and reliability of the French-language version of the positive and negative syndrome scale (PANSS). *Acta Psychiatr Scand*. 1999;100(3):237–243.
- 75 Bunk D, Eggers C, Klapal M. Symptom dimensions in the course of childhood-onset schizophrenia. *Eur Child Adolesc Psychiatry*. 1999;8(Suppl 1):129–135.
- 76 Lançon C, Aghababian V, Llorca PM, Auquier P. Factorial structure of the Positive and Negative Syndrome Scale (PANSS): a forced five-dimensional factor analysis. *Acta Psychiatr Scand*. 1998;98(5):369–376.
- 77 Klapal M, Eggers C, Bunk D, Koriath H. [The 5 factor model of childhood schizophrenia]. *Nervenarzt*. 1998;69(3):238–242.
- 78 Higashima M, Urata K, Kawasaki Y, et al. P300 and the thought disorder factor extracted by factor-analytic procedures in schizophrenia. *Biol Psychiatry*. 1998;44(2):115–120.
- 79 Loas G, Noisette C, Legrand A, Delahousse J. A four-syndrome model of chronic schizophrenia: principal components analysis of the Positive and Negative Syndrome Scale (PANSS) in 153 chronic schizophrenics. *Encephale*. 1997;23(1):10–18.
- 80 Lançon C, Auquier P, Llorca PM, Martínez JL, Bougerol T, Scotto JC. Psychometric properties of the PANSS in a sample of French schizophrenic patients. *Encephale*. 1997;23(1):1–9.
- 81 Dollfus S, Petit M. Principal-component analyses of PANSS and SANS-SAPS in schizophrenia: their stability in an acute phase. *Eur Psychiatry*. 1995;10(2):97–106.
- 82 Peralta V. Psychometric properties of the positive and negative syndrome scale (PANSS) in schizophrenia. *Psychiatry Res*. 1994;53(1):31–40.
- 83 Lindenmayer JP, Bernstein-Hyman R, Grochowski S. A new five factor model of schizophrenia. *Psychiatr Q*. 1994;65(4):299–322.
- 84 Lindenmayer JP, Bernstein-Hyman R, Grochowski S. Five-factor model of schizophrenia. Initial validation. *J Nerv Ment Dis*. 1994;182(11):631–638.
- 85 Kawasaki Y, Maeda Y, Sakai N, et al. Evaluation and interpretation of symptom structures in patients with schizophrenia. *Acta Psychiatr Scand*. 1994;89(6):399–404.
- 86 Bell MD. Five-component model of schizophrenia: assessing the factorial invariance of the positive and negative syndrome scale. *Psychiatry Res*. 1994;52(3):295–303.
- 87 Lindström E, von Knorring L. Principal component analysis of the Swedish version of the positive and negative syndrome scale for schizophrenia. *Nord J Psychiatr*. 1993;47(4):257–263.
- 88 Lepine J. Dimensions positives et négatives dans les schizophrénies. *Les Cahiers de Prisme*. 1991;1:23–29.
- 89 Dollfus S, Petit M, Lesieur P, Menard JF. Principal-component analysis of PANSS and SANS-SAPS global ratings in schizophrenic patients. *Eur Psychiatry*. 1991;6(5):251–259.
- 90 Kay SR, Sevy S. Pyramidal model of schizophrenia. *Schizophr Bull*. 1990;16(3):537–545.
- 91 Baandrup L, Allerup P, Nielsen MO, et al. Rasch analysis of the PANSS negative subscale and exploration of negative symptom trajectories in first-episode schizophrenia - data from the OPTi-MiSE trial. *Psychiatry Res*. 2020;289:112970.
- 92 Cruz BF, de Resende CB, Abreu MN, et al. How specific are negative symptoms and cognitive impairment in schizophrenia? An analysis of PANSS and SCoRS. *Cogn Neuropsychiatry*. 2013;18(3):243–251.
- 93 Lindström E, Wieselgren IM, von Knorring L. Interrater reliability of the structured clinical interview for the positive and negative syndrome scale for schizophrenia. *Acta Psychiatr Scand*. 1994;89(3):192–195.
- 94 Barrio C, Yamada AM, Atuel H, et al. A tri-ethnic examination of symptom expression on the positive and negative syndrome scale in schizophrenia spectrum disorders. *Schizophr Res*. 2003;60(2–3):259–269.
- 95 Ivanova E, Khan A, Liharska L, et al. Validation of the Russian version of the positive and negative syndrome scale (PANSS-Ru) and normative data. *Innov Clin Neurosci*. 2018;15(9–10):32–48.
- 96 Hashimoto N, Takahashi K, Fujisawa D, et al. A pilot validation study of the Japanese translation of the Positive and Negative Syndrome Scale (PANSS). *Asian J Psychiatr*. 2020;54:102210.
- 97 Li L, Ma H, Wang X, Meng E. Validation of Chinese version of positive and negative syndrome scale-6 in clinical setting: a preliminary study. *Psychiatry Clin Psychopharmacol*. 2021;31(4):386–391.
- 98 Medvedev ON, Berk M, Dean OM, et al. A novel way to quantify schizophrenia symptoms in clinical trials. *Eur J Clin Invest*. 2021;51(3):e13398.
- 99 Kay SR, Fiszbein A, Lindenmayer JP, Opler LA. Positive and negative syndromes in schizophrenia as a function of chronicity. *Acta Psychiatr Scand*. 1986;74(5):507–518.
- 100 Kay SR, Opler LA, Fiszbein A. Significance of positive and negative syndromes in chronic schizophrenia. *Br J Psychiatry*. 1986;149:439–448.
- 101 Bell M, Milstein R, Beam-Goulet J, Lysaker P, Cicchetti D. The positive and negative syndrome scale and the brief psychiatric rating scale. Reliability, comparability, and predictive validity. *J Nerv Ment Dis*. 1992;180(11):723–728.

- 102 von Knorring L, Lindström E. The Swedish version of the Positive and Negative Syndrome Scale (PANSS) for schizophrenia. Construct validity and interrater reliability. *Acta Psychiatr Scand*. 1992;86(6):463–468.
- 103 Peralta V, Cuesta MJ, de Leon J. Positive and negative symptoms/syndromes in schizophrenia: reliability and validity of different diagnostic systems. *Psychol Med*. 1995;25(1):43–50.
- 104 Vollema MG, Geurtsen GJ, Kuipers T. Negative symptoms: a uni-dimensional construct related to frontal lobe deficits. *Tijdschr Psychiatr*. 1995;37(8):615–628.
- 105 Norman RM, Malla AK, Cortese L, Diaz F. A study of the interrelationship between and comparative interrater reliability of the SAPS, SANS and PANSS. *Schizophr Res*. 1996;19(1):73–85.
- 106 Igarashi Y, Hayashi N, Yamashina M, et al. Interrater reliability of the Japanese version of the Positive and Negative Syndrome Scale and the appraisal of its training effect. *Psychiatry Clin Neurosci*. 1998;52(5):467–470.
- 107 Anil Yağcıoğlu AE, Batur S, Tiryaki A, Gogus A. Reliability and validity of the Turkish version of the positive and negative syndrome scale (PANSS). *Türk Psikol Derg*. 1999;14:23–34.
- 108 Chen J, Patil KR, Weis S, et al. Neurobiological divergence of the positive and negative schizophrenia subtypes identified on a new factor structure of psychopathology using non-negative factorization: an international machine learning study. *Biol Psychiatry*. 2020;87(3):282–293.
- 109 Kay SR, Fiszbein A, Vital-Herne M, Fuentes LS. The positive and negative syndrome scale - Spanish adaptation. *J Nerv Ment Dis*. 1990;178(8):510–517.
- 110 Nilchaikovit T, Ueanong S, Kessawai D, Thomyangkoon P. The Thai version of the Positive and Negative Syndrome Scale (PANSS) for schizophrenia: criterion validity and interrater reliability. *J Med Assoc Thai*. 2000;83(6):646–651.
- 111 Aggarwal NK, Tao H, Xu K, Stefanovics E, Zhening L, Rosenheck RA. Comparing the PANSS in Chinese and American inpatients: cross-cultural psychiatric analyses of instrument translation and implementation. *Schizophr Res*. 2011;132(2–3):146–152.
- 112 Aggarwal NK, Zhang XY, Stefanovics E, et al. Rater evaluations for psychiatric instruments and cultural differences: the positive and negative syndrome scale in China and the United States. *J Nerv Ment Dis*. 2012;200(9):814–820.
- 113 Bentsen H, Munkvold OG, Notland TH, et al. The interrater reliability of the positive and negative syndrome scale (PANSS). *Int J Methods Psychiatr Res*. 1996;6:227–235.
- 114 Müller MJ, Rossbach W, Dannigkei P, Müller-Siecheneder F, Szegedi A, Wetzel H. Evaluation of standardized rater training for the positive and negative syndrome scale (PANSS). *Schizophr Res*. 1998;32(3):151–160.
- 115 Müller MJ, Wetzel H. Improvement of inter-rater reliability of PANSS items and subscales by a standardized rater training. *Acta Psychiatr Scand*. 1998;98(2):135–139.
- 116 Khan A, Yavorsky WC, Liechti S, et al. Assessing the sources of unreliability (rater, subject, time-point) in a failed clinical trial using items of the Positive and Negative Syndrome Scale (PANSS). *J Clin Psychopharmacol*. 2013;33(1):109–117.
- 117 Kay SR, Opler LA, Lindenmayer JP. Reliability and validity of the positive and negative syndrome scale for schizophrenics. *Psychiatry Res*. 1988;23(1):99–110.
- 118 Jelastopulu E, Giourou E, Merekoulis G, Mestousi A, Moratis E, Alexopoulos EEC. Correlation between the personal and social performance scale (PSP) and the positive and negative syndrome scale (PANSS) in a Greek sample of patients with schizophrenia. *BMC Psychiatry*. 2013;14(1):197.
- 119 Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR. What does the PANSS mean? *Schizophr Res*. 2005;79(2–3):231–238.
- 120 Edgar CJ, Blaettler T, Bugarski-Kirola D, Le Scouiller S, Garibaldi GM, Marder SR. Reliability, validity and ability to detect change of the PANSS negative symptom factor score in outpatients with schizophrenia on select antipsychotics and with prominent negative or disorganized thought symptoms. *Psychiatry Res*. 2014;218(1–2):219–224.
- 121 Callegari C, Poloni N, Pace V, et al. Evaluation of positive and negative symptoms in chronic schizophrenia: an interrelationship study among SANS/SAPS, PANSS and InSka rating scales. *Minerva Psychiatr*. 2005;46(4):315–323.
- 122 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;137(1–3):246–250.
- 123 Van Erp TG, Preda A, Nguyen D, et al. Converting positive and negative symptom scores between PANSS and SAPS/SANS. *Schizophr Res*. 2014;152(1):289–294.
- 124 Hieronymus F, Correll CU, Østergaard SD. Initial severity of the Positive and Negative Syndrome Scale (PANSS)-30, its main subscales plus the PANSS-6, and the relationship to subsequent improvement and trial dropout: a pooled participant-level analysis of 18 placebo-controlled risperidone and paliperidone trials. *Transl Psychiatry*. 2023;13(1):191.
- 125 Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR. Linking the PANSS, BPRS, and CGI: clinical implications. *Neuropsychopharmacology*. 2006;31(10):2318–2325.
- 126 Leucht S, Rothe P, Davis JM, Engel RR. Equipercentile linking of the BPRS and the PANSS. *Eur Neuropsychopharmacol*. 2013;23(8):956–959.
- 127 Hieronymus F, Kølbaek P, Correll CU, Østergaard SD. Antipsychotic-placebo separation on the PANSS-6 subscale as compared to the PANSS-30: a pooled participant-level analysis. *NPJ Schizophr*. 2021;7(1):41.
- 128 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.
- 129 Fitzgerald PB, de Castella AR, Brewer K, et al. A confirmatory factor analytic evaluation of the pentagonal PANSS model. *Schizophr Res*. 2003;61(1):97–104.
- 130 Langeveld J, Andreassen OA, Auestad B, et al. Is there an optimal factor structure of the Positive and Negative Syndrome Scale in patients with first-episode psychosis? *Scand J Psychol*. 2013;54(2):160–165.
- 131 Rodriguez-Jimenez R, Bagney A, Mezquita L, et al. Cognition and the five-factor model of the positive and negative syndrome scale in schizophrenia. *Schizophr Res*. 2013;143(1):77–83.
- 132 Stefanovics EA, Elks H, Zhening L, Zhang XY, Rosenheck RA. A cross-national factor analytic comparison of three models of PANSS symptoms in schizophrenia. *Psychiatry Res*. 2014;219(2):283–289.
- 133 Woodward TS, Jung K, Smith GN, et al. Symptom changes in five dimensions of the Positive and Negative Syndrome Scale in refractory psychosis. *Eur Arch Psychiatr Clin Neurosci*. 2014;264(8):673–682.
- 134 Fong TC, Ho RT, Wan AH, Siu PJ, Au-Yeung FS. Psychometric validation of the consensus five-factor model of the positive and negative syndrome scale. *Compr Psychiatry*. 2015;62:204–208.
- 135 Bentler PM. Comparative fit indexes in structural models. *Psychol Bull*. 1990;107(2):238–246.
- 136 Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res Sci Educ*. 2018;48(6):1273–1296.
- 137 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297–334.
- 138 Ratner B. The correlation coefficient: its values range between +1/–1, or do they? *J Target Meas Anal Market*. 2009;17(2):139–142.
- 139 O'Brien RM. Generalizability coefficients are reliability coefficients. *Qual Quantity*. 1995;29(4):421–428.
- 140 Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Med Teach*. 2012;34(11):960–992.
- 141 Kay SR, Opler LA, Lindenmayer J. *Structured Clinical Interview for the PANSS (SCI-PANSS) Multi-Health Systems*. 1992.
- 142 Kay SR, Opler LA, Fiszbein A. *Positive and Negative Syndrome Scale (PANSS): Manual*. Toronto, ON: Multi-Health Systems, Inc; 2006.
- 143 Khan A, Liharska L, Harvey PD, Atkins A, Ulshen D, Keefe RSE. Negative symptom dimensions of the positive and negative syndrome scale across geographical regions: implications for social, linguistic, and cultural consistency. *Innov Clin Neurosci*. 2017;14(11–12):30–40.
- 144 Pearson Assessments. Positive and negative syndrome scale. <https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Personality-%26-Biopsychosocial/Positive-and-Negative-Syndrome-Scale/p/P100025000.html>; 2023. Accessed August 23, 2024.
- 145 European Medicines Agency. *Guideline on Clinical Investigation of Medicinal Products, Including Depot Preparations, in the Treatment of*

- Schizophrenia - Revision 1*. Amsterdam, The Netherlands: European Medicines Agency; 2012.
- 146 Leucht S, Engel RR. The relative sensitivity of the clinical global impressions scale and the brief psychiatric rating scale in antipsychotic drug trials. *Neuropsychopharmacology*. 2006;31(2):406–412.
 - 147 Sifakis S, Brandt L, McCutcheon RA, et al. Relapse in clinically stable adult patients with schizophrenia or schizoaffective disorder: evidence-based criteria derived by equipercentile linking and diagnostic test accuracy meta-analysis. *Lancet Psychiatry*. 2024;11(1):36–46.
 - 148 Leucht S, Barabási Á, Laszlovszky I, et al. Linking PANSS negative symptom scores with the Clinical Global Impressions Scale: understanding negative symptom scores in schizophrenia. *Neuropsychopharmacology*. 2019;44(9):1589–1596.
 - 149 Levine SZ, Rabinowitz J, Engel R, Etschel E, Leucht S. Extrapolation between measures of symptom severity and change: an examination of the PANSS and CGI. *Schizophr Res*. 2008;98(1–3):318–322.
 - 150 Leucht S, Davis JM, Engel RR, Kane JM, Wagenpfeil S. Defining ‘response’ in antipsychotic drug trials: recommendations for the use of scale-derived cutoffs. *Neuropsychopharmacology*. 2007;32(9):1903–1910.
 - 151 Leucht S, Davis JM, Engel RR, Kissling W, Kane JM. Definitions of response and remission in schizophrenia: recommendations for their use and their presentation. *Acta Psychiatr Scand Suppl*. 2009;438:7–14.
 - 152 Obermeier M, Mayr A, Schennach-Wolff R, Seemüller F, Möller HJ, Riedel M. Should the PANSS be rescaled? *Schizophr Bull*. 2010;36(3):455–460.
 - 153 The PANSS Institute. PANSS.org. <https://panss.org/services-training.php>; 2024. Accessed August 22, 2024.
 - 154 Marder SR, Kirkpatrick B. Defining and measuring negative symptoms of schizophrenia in clinical trials. *Eur Neuropsychopharmacol*. 2014;24(5):737–743.
 - 155 Kring AM, Gur RE, Blanchard JJ, Horan WP, Reise SP. The clinical assessment interview for negative symptoms (CAINS): final development and validation. *Am J Psychiatry*. 2013;170(2):165–172.
 - 156 Kirkpatrick B, Strauss GP, Nguyen L, et al. The brief negative symptom scale: psychometric properties. *Schizophr Bull*. 2011;37(2):300–305.
 - 157 Horan WP, Kring AM, Gur RE, Reise SP, Blanchard JJ. Development and psychometric validation of the clinical assessment interview for negative symptoms (CAINS). *Schizophr Res*. 2011;132(2–3):140–145.
 - 158 Forbes C, Blanchard JJ, Bennett M, Horan WP, Kring A, Gur R. Initial development and preliminary validation of a new negative symptom measure: the Clinical Assessment Interview for Negative Symptoms (CAINS). *Schizophr Res*. 2010;124(1–3):36–42.
 - 159 Alphs LD, Summerfelt A, Lann H, Muller RJ. The negative symptom assessment: a new instrument to assess negative symptoms of schizophrenia. *Psychopharmacol Bull*. 1989;25(2):159–163.
 - 160 Andreasen NC. *The scale for the assessment of negative symptoms (SANS)*. Iowa City, IA: University of Iowa; 1983.
 - 161 Németh G, Laszlovszky I, Czobor P, et al. Cariprazine versus risperidone monotherapy for treatment of predominant negative symptoms in patients with schizophrenia: a randomised, double-blind, controlled trial. *Lancet*. 2017;389(10074):1103–1113.
 - 162 Leucht S, Leucht C, Huhn M, et al. Sixty years of placebo-controlled antipsychotic drug trials in acute schizophrenia: systematic review, Bayesian meta-analysis, and meta-regression of efficacy predictors. *Am J Psychiatry*. 2017;174(10):927–942.
 - 163 Lindenmayer JP. Are shorter versions of the positive and negative syndrome scale (PANSS) doable? A critical review. *Innov Clin Neurosci*. 2017;14(11–12):73–76.
 - 164 Yamamoto N, Inada T, Shimodera S, Morokuma I, Furukawa TA. Brief PANSS to assess and monitor the overall severity of schizophrenia. *Psychiatr Clin Neurosci*. 2010;64(3):262–267.
 - 165 Andreasen NC, William T, Carpenter J, et al. Remission in schizophrenia: proposed criteria and rationale for consensus. *Am J Psychiatry*. 2005;162(3):441–449.
 - 166 Takeuchi H, Lee J, Fervaha G, Agid O, Remington G. Second version of brief evaluation of psychosis symptom domains (BE-PSD-V2.0). *Schizophr Res*. 2023;254:176–177.
 - 167 Geck S, Roithmeier M, Bühner M, et al. COSMIN systematic review and meta-analysis of the measurement properties of the PANSS-6. *Eur Neuropsychopharmacol*. 2025;94:41–50.
 - 168 Esfahlani FZ, Sayama H, Visser KF, Strauss GP. Sensitivity of the positive and negative syndrome scale (PANSS) in detecting treatment effects via network analysis. *Innov Clin Neurosci*. 2017;14(11–12):59–67.
 - 169 Kalkbrenner MT. Alpha, omega, and H internal consistency reliability estimates: reviewing these options and when to use them. *Counsel Outcome Res Eval*. 2023;14(1):77–88.
 - 170 Fabrigar LR, Wegener DT, MacCallum R, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods*. 1999;4:272–299.
 - 171 Huhn M, Nikolakopoulou A, Schneider-Thoma J, et al. Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multi-episode schizophrenia: a systematic review and network meta-analysis. *Lancet*. 2019;394(10202):939–951.
 - 172 Takeuchi H, Fervaha G, Lee J, Agid O, Remington G. A preliminary examination of the validity and reliability of a new brief rating scale for symptom domains of psychosis: brief Evaluation of Psychosis Symptom Domains (BE-PSD). *J Psychiatr Res*. 2016;80:87–92.
 - 173 Haro JM, Karmath SA, Ochoa S, et al. The Clinical Global Impression-Schizophrenia scale: a simple instrument to measure the diversity of symptoms present in schizophrenia. *Acta Psychiatr Scand Suppl*. 2003;416:16–23.