Schizophrenia Research 240 (2022) 32–45

Contents lists available at ScienceDirect

Schizophrenia Research

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

Reliability of the Calgary depression scale for schizophrenia: A meta-analysis

Laura Porter *, Christopher Jones, Andrew Fox

Centre for Applied Psychology at the University of Birmingham, United Kingdom of Great Britain and Northern Ireland
Birmingham and Solihull Mental Health Foundation NHS Trust, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Keywords:
Internal consistency
Inter-rater reliability
Negative symptoms
Depression
Review

ABSTRACT

Background: A challenge for clinicians working with individuals diagnosed with schizophrenia is distinguishing depressive symptoms from negative symptoms of schizophrenia. The Calgary Depression Scale for Schizophrenia (CDSS) was developed for this purpose. No review has previously explored its reliability across multiple studies using advanced statistical means.

Objectives: This meta-analysis aimed to quantify the CDSS' internal consistency, inter-rater reliability (IRR) and test-retest reliability.

Method: A systematic literature search was conducted to find articles reporting on the CDSS' reliability. Articles were screened against the inclusion and exclusion criteria, with data extracted from 40 studies. Overall meta-analytic effects were calculated, and for internal consistency and IRR coefficients subsequent analyses explored between-study variation. The small test-retest reliability dataset limited analysis.

Findings: The internal consistency meta-analytic effect was 0.83 (95% CI:0.82–0.84). Higgins I^2 indicated an acceptable level of variation between studies' alpha estimates. This suggests all items in the CDSS are measuring the same construct (i.e. symptoms of depression). The IRR meta-analytic effect was 0.88 (95% CI:0.86–0.91), with Higgins I^2 indicating high levels of heterogeneity. This was not deemed problematic variance as it is within levels expected for psychometric measures and, therefore, considered acceptable for this literature. This reflects high level of agreement between different raters when using the CDSS on the same client.

Conclusions: This review suggests the CDSS has good internal consistency and excellent IRR. Further research will help understand its test-retest reliability.

1. Introduction

Schizophrenia is associated with a range of symptoms, typically separated into positive and negative. Positive symptoms are associated with the individual's perception or interpretation of stimuli being different from others, alongside difficulties distinguishing their thoughts and ideas from reality. Negative symptoms include loss of motivation, apathy, impaired concentration, flattening of emotions and reduced speech (Cuesta et al., 2009).

It is widely accepted that mood disturbances are often observed alongside a diagnosis of schizophrenia (Rector et al., 2005; van Os et al., 2000). This includes mood disturbances experienced concurrently and independently from the psychotic symptoms (Birchwood et al., 2000). Negative symptoms, however, overlap with symptoms of depression, posing a clinical challenge of distinguishing negative symptoms of schizophrenia (i.e. difficulties with motivational state) from depression (i.e. difficulties with pervasive low mood).

A recent meta-analytic review found an association between higher rates of negative symptoms and higher rates of depressive symptoms (Edwards et al., 2019). There is evidence that the presence of depressive symptoms is associated with a poorer prognosis and an increased prevalence of suicide, compared to those diagnosed with schizophrenia not displaying depressive symptoms (Buckley et al., 2009; Conley et al., 2007; Upthegrove et al., 2016). Negative symptoms have also been associated with poorer functional outcomes in those diagnosed with schizophrenia (Foussias and Remington, 2010). It is, therefore, important to be able to identify signs of affective dysregulation (e.g. low mood) in individuals diagnosed with schizophrenia. This enables clinicians to distinguish between extrapyramidal, negative, and depressive symptoms.

* Corresponding author at: Pritchatts Road, University of Birmingham, Birmingham B15 2SA, United Kingdom of Great Britain and Northern Ireland.
E-mail address: lsp759@alumni.bham.ac.uk (L. Porter).

https://doi.org/10.1016/j.schres.2021.11.040
Received 27 February 2021; Received in revised form 17 October 2021; Accepted 26 November 2021
Available online 14 December 2021

The Calgary Depression Scale for Schizophrenia was developed as a specific measure of depression in individuals diagnosed with schizophrenia (Addington et al., 1990). The CDSS assesses the level of depression in such individuals. It distinguishes between depression and the positive and negative symptoms present in schizophrenia (Addington et al., 1994). The CDSS is suitable for individuals in the acute and residual stages of schizophrenia and is sensitive to change in presentation over time (Addington et al., 1993). The 9-item scale is completed by clinicians experienced in working with individuals diagnosed with schizophrenia. Items 1–8 are rated via interview to assess the presence of symptoms over the past two weeks. The final item's rating is dependent on the interviewer's observations throughout the interview (Addington et al., 1993). Since its development, the CDSS has been translated into different languages, with 44 language variants existing.

Reliability of a measure refers to its level of consistency, with different types existing (Price et al., 2015). A measure's reliability is important to know to establish the suitability of the tool for clinical and research purposes. Measures with poor reliability may produce unstable or confounded results.

The CDSS is routinely used in clinical practice and research. Clinically it is used to identify individuals at increased risk of attempting suicide and those requiring intervention to address their symptoms of depression (Addington et al., 1993). NICE guidelines recommend individuals diagnosed with schizophrenia are routinely assessed for depression (BPS, 2017; NICE, 2014). It is, therefore, important to understand the CDSS's reliability to establish its suitability.

The reliability of the CDSS has been considered as part of a systemic review (Lako et al., 2012). Lako et al.'s review (2012) considered the use of six measures of depression with individuals diagnosed with schizophrenia. The CDSS was reported to have good reliability and validity, when compared to the other tools. The authors recommended the CDSS for clinical practice and research. This review combined results by taking a mean of the reliability values across individual studies rather than more advanced analytical means. Meta-analysis provides a means of synthesising quantitative findings from different studies using statistical methods (Rodriguez and Maeda, 2006). This approach considers individual study results together, overcoming limitations associated with smaller sample sizes and individual study biases (Walker et al., 2008). This is important for measures, like the CDSS, that have been translated into different languages and used with different diagnoses (Rodriguez and Maeda, 2006). There is currently no review providing a detailed numerical meta-analytic synthesis and comparison of the CDSS' psychometric properties.

The overall aim of this meta-analysis is to investigate the CDSS' reliability. Consideration will be given to the version of the CDSS used and the sample's diagnostic composition. This is due to the CDSS having been translated into multiple languages and having been developed specifically for individuals diagnosed with schizophrenia.

2. Methods

2.1. Search strategy

A literature search of three databases (PsycINFO, Medline and PubMed) was conducted between May 2019 and October 2021. Reference lists and articles cited in full-text articles were reviewed to identify additional studies, alongside reviewing the articles included in Lako et al.'s review (2012). The aim was to identify literature reporting the CDSS' reliability to assess the measure's reliability across studies. The following search terms were used: {"Calgary Depression Scale" or CDSS} AND {reliability OR "internal consistency" OR alpha OR test-retest} AND {schizophrenia OR schizophrenic}.

2.2. Eligibility criteria

The initial search of the databases identified 1016 records. A further 35 records were found through searching references of articles from the database search. After removing duplicate records, the remaining 950 records were screened for eligibility by applying the inclusion and exclusion criteria (Table 1). The literature search and the search outcomes are presented in Fig. 1, alongside an explanation for each criterion.

2.3. Data extraction

The process of data extraction was completed by the lead author (LP). Any queries in relation to the process of extracting data, such as whether it met the inclusion criteria or how to convert it to an appropriate form (e.g. a reliability coefficient), were discussed with CJ and AP to make a decision by consensus.

From each paper, descriptive data was extracted regarding author, study setting, sample size, sample demographics (e.g. age, gender, diagnosis and onset of illness) and CDSS version used (i.e. original or translated). Data were also extracted for synthesis, including reliability coefficients and number of items. For studies reporting IRR coefficients, the type of coefficient (e.g. Kappa or ICC) was noted. Regarding IRR, of the included studies, two reported the % agreement between raters. For these two studies (Mingrone et al., 2016; Pawełczyk et al., 2016), this value was converted into a Kappa coefficient using the following formula (Glen, 2014):

$$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}$$

where:

P_o = the relative observed agreement among raters.
P_e = the hypothetical probability of chance agreement.

Table 1
Inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria	Rationale for criteria
Articles published in the English language between 1990 and October 2021.	Article published in a foreign language (i.e. not in English).	This is due to time and resource constraints preventing non-English papers being translated.
Original articles reporting original empirical data	Review articles, study protocols, critique papers and books/book chapters. Journal articles reporting reliability estimates from previous studies (i.e. not original data values).	The focus of this review is on studies that present original data, which is required for inclusion within a meta-analysis. Excluded formats reflect those that do not provide such data.
Study sample taken from a population reflecting the intended population for which the CDSS was designed.	The main diagnosis (i.e. <50%) within the sample was not schizophrenia.	The CDSS was developed for use specifically with individuals who have a diagnosis of schizophrenia. The sample was, therefore, required to have at least 50% of participants with a diagnosis of schizophrenia.
Use of the CDSS in its original form or a translated version which has been approved by the scale developer.	The CDSS, in either the original or a translated form, was not used.	The focus of this meta-analysis is on the CDSS. As such, only studies using an approved version of this measure were included.
Appropriate statistical reporting of at least one type of reliability for the CDSS.	Primary studies in which no reliability coefficients for the CDSS were reported.	The aim of this meta-analysis is to investigate the reliability of the CDSS. As such, appropriate reporting of reliability coefficients is a requirement for the study to be included within the analysis.

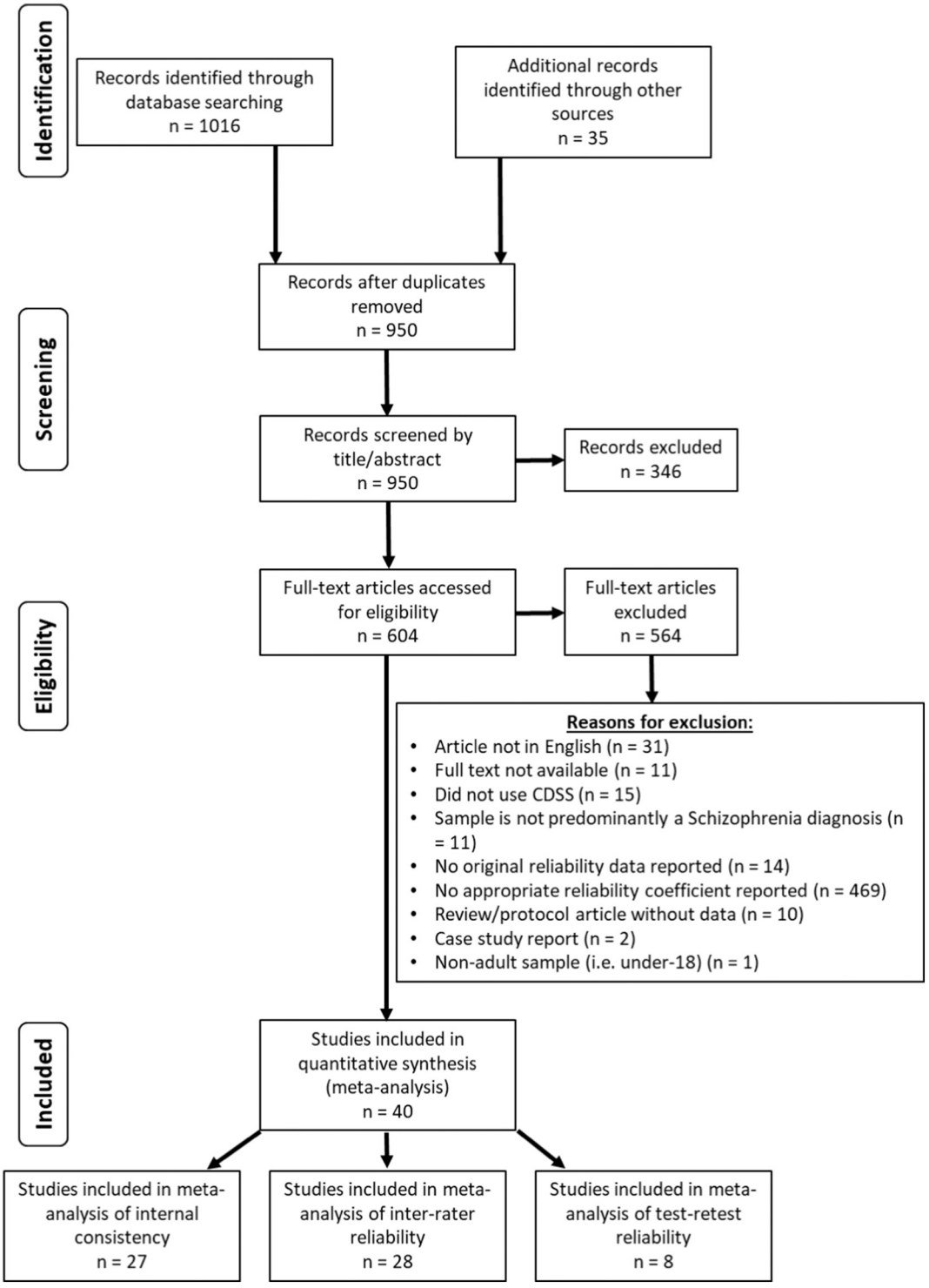

Fig. 1. Flow chart of the search strategy and process of article selection.[a]
[a]Based on Moher, Liberati, Tetzlaff, Altman & The PRISMA Group's diagram (Moher et al., 2009).

There were two studies (Kaneda et al., 2000; Xiao et al., 2009) where individual item ICC values were reported but an overall ICC was not provided. In these cases, an overall value was calculated by taking the average of Fisher's z transformed kappa values.

Before numerical synthesis was computed, alpha estimates and IRR coefficients from the studies were transformed using Fisher's (Fisher, 1921) z transformation. This transformation is recommended as it functions to normalise the distribution of effects and stabilises the variance of the estimates due to the non-linearity of correlational data (Corey et al., 1998). Following synthesis, the Fisher's z scores were transformed back to Pearson's R values for ease of reporting.

More details of the data analysis strategy are provided in supplementary appendix 1.

3. Results

3.1. Study characteristics

Table 2 presents the characteristics of the 39 studies included in the meta-analysis. Of the 40 studies, 27 were included for the CDSS' internal consistency and 28 were included for the IRR of the CDSS. Eight of these studies also reported test-retest reliability data, however, due to limitations within this dataset the analysis was restricted (further information in Sections 3 and 4).

3.2. Risk of bias of individual studies

Findings of meta-analyses can be impacted by including poor quality studies (Higgins et al., 2011). Quality captures how appropriate the study is for answering its research question, considering design, delivery and analysis. There are various tools for assessing risk of bias. Higgins et al. (2011) advocate using a set of criteria specific to methodological issues pertinent to the literature and question under review. Assessment of risk of bias, therefore, was completed using a framework developed for this review (Table 3). Existing tools and information on types of bias guided the framework's development (Higgins et al., 2011; Smith and Noble, 2014).

Risk of bias ratings were primarily made by the lead author (LP), who also completed the process of data extraction. When there was ambiguity within the data which complicated the decision as to whether there was a low, unclear or high risk of bias, a discussion was had between the authors (LP, CJ and AF) to enable the final decision to be made by consensus.

3.3. Review of risk of bias

3.3.1. Selection bias

Selection bias was mixed, with 25 rated low risk and 15 unclear risk. For those rated unclear, this was due to including participants not diagnosed with schizophrenia. These were papers where the sample contained a mixture of psychotic diagnoses whereby less than 100% but more than 50% of the sample were diagnosed with schizophrenia (i.e. the diagnosis for which the CDSS is designed).

3.3.2. Performance bias

There was variation in performance bias across the studies, with 25 rated low risk, 14 unclear risk and one high risk. Krupchanka and Katliar (2016) was rated high risk due to raters not having been trained to use the measure, opposing the CDSS' protocol, and it being unclear whether the translated version had been approved by the scale's author. For the 14 studies considered unclear risk, this was because it was unclear whether raters were trained according to Addington et al.'s (1990) protocol.

3.3.3. Reporting bias

Reporting bias was generally low within the primary studies, with 37 rated low risk, two unclear risk and one high risk. Schuetze et al. (2001) was considered high risk given data reported on the measures used was limited. For example, no participant scores were reported for each measure, such as mean, with only psychometric properties data provided. The two studies deemed unclear risk were due to partial reporting of descriptive statistics (Bull et al., 2016; Müller et al., 1999). In Müller et al.'s (1999) study, there was limited demographic information (only age and diagnosis) and CDSS scores were not reported. Bull et al. (2016) indicated participants were separated into two groups, but the number per group was not reported, impacting data interpretation.

3.3.4. Detection bias

Depending on the type of reliability, ratings of detection bias varied across the studies. For IRR and test-retest reliability, all primary studies were rated low risk. This was due to clear reporting on reliability coefficients of the measure used as applied to their study participants. One exception to this was (Rostami et al., 2019) which was rated unclear due to IRR not being reported. For internal consistency, detection bias was mixed, with 28 rated low risk and 12 unclear risk. For the 12 studies considered to have an unclear risk, this was due to authors not reporting the IRR of the CDSS for their study.

3.3.5. Statistical bias[1]

Depending on the reliability being assessed, there was some variation in statistical bias ratings. For all studies reporting internal consistency coefficients (N = 27), statistical bias was considered low risk. Within these studies, 15 also provided IRR coefficients. Of these, six were rated unclear risk for statistical bias (Coulston et al., 2007; Kaneda et al., 2000*; Kontaxakis et al., 2000*; Schennach et al., 2012; Xiao et al., 2009*; Xu et al., 2018). Unclear risk for IRR data was due to non-exact reliability coefficients being reported and/or uncertainty over the percentage of the population used to calculate IRR. For one paper no overall IRR was reported, meaning it had to be calculated from individual item reliability coefficients (Xiao et al., 2009). Of the studies only reporting IRR coefficients (N = 13), nine were considered unclear risk of statistical bias (Chengappa et al., 2012; Fathian et al., 2019; Fitzgerald et al., 2008; Jäger et al., 2008; Jeon et al., 2018; Lincoln et al., 2010; Müller et al., 2006; Pawełczyk et al., 2016; Zisook et al., 2006). These ratings related to uncertainty about how IRR had been calculated (i.e. on a subset of the sample or the whole sample).

3.3.6. Generalisability

Depending on the nature of the reliability being assessed, ratings of generalisability varied across the studies. Regarding the internal consistency dataset, 24 studies were considered low risk of generalisability and three (Kontaxakis et al., 2000; Schuetze et al., 2001; Xiao et al., 2009) unclear risk due to having sample sizes between 20 and 40 participants.

For the IRR data, there was greater variation in risk ratings for generalisability. There were 12 studies rated low risk, nine studies unclear risk and seven studies high risk of generalisability. Unclear risk was assigned to studies using a sample of 20–40 participants (Bernard et al., 1998; Fitzgerald et al., 2008; Hani et al., 2016; Kontaxakis et al., 2000; Maggini and Raballo, 2004; Müller et al., 2006; Sarró et al., 2004; Schuetze et al., 2001; Xiao et al., 2009). High risk studies were those with a sample size less than 20 participants, accounting for seven studies (Addington et al., 1992, 1994, 1996; Kaneda et al., 2000; Mingrone et al., 2016; Müller et al., 1999; Suttajit et al., 2013).

For the test-retest reliability data, risk ratings for generalisability were more varied. Two studies were rated low risk, three unclear risk and three high risk of generalisability. Studies were assigned unclear risk rating due to having sample sizes between 20 and 40 (Kaneda et al., 2000; Kontaxakis et al., 2000; Xiao et al., 2009). High risk ratings were for those studies with samples smaller than 20 (Bernard et al., 1998; Hani et al., 2016; Sarró et al., 2004).

3.4. Internal consistency

There were 27 studies reporting alpha coefficients in a total sample size of 3024 participants. Sample sizes in the primary studies ranged from 20 (Schuetze et al., 2001) to 349 (Xu et al., 2018). The distribution of the study level estimates of Fisher's (Fisher, 1921) transformation of internal reliability coefficients conformed to normal expectations and the DerSimonian-Laird (DerSimonian and Laird, 1986) method of calculating heterogeneity was considered appropriate, as shown in the QQ plot (Fig. 2). The DerSimonian-Laird estimate is, therefore,

[1] Studies marked with an * reflect studies which were also assigned an unclear risk rating for the test-retest reliability data for the same reason.

Table 2
Overview of included studies.[a]

Study	Year	Internal consistency	Inter-rater reliability (N)	Test-retest reliability (N)	N	Sample diagnosis	Sample Type	Version of the CDSS	Country in which study conducted	Context of CDSS use
Addington et al.	1992	0.79	0.90 (10)		150	100% schizophrenia	Inpatients & Outpatients	Original	Canada	Reliability and validity of CDSS
Addington et al.	1994	0.84	0.96 (10)		150	100% schizophrenia	Inpatients & Outpatients	Original	Canada	Establishing CDSS's specificity
Addington et al.	1996	0.82	0.89 (10)		112	100% schizophrenia	Inpatients & Outpatients	Original	Canada	Comparison of CDSS with another measure
Bernard et al.	1998	0.79	0.98 (33)	0.69 (16)	70	100% schizophrenia	Inpatients & Outpatients	Translated	France	Validation of CDSS
Bressan et al.	1998	0.8			80	100% schizophrenia	Outpatients	Translated	Brazil	Validation of CDSS
Lançon et al.	1999	0.82			95	100% schizophrenia	Inpatients & Outpatients	Translated	France	Reliability and validity of CDSS
Müller et al.	1999		0.97 (10)		10	Mixed psychotic disorders	Inpatients	Translated	Germany	Reliability of CDSS
Kaneda et al.	2000	0.82	0.84 (11)	0.86 (28)	47	100% schizophrenia	Inpatients	Translated	Japan	Validation of CDSS
Kontaxakis et al.	2000	0.87	0.78 (24)	0.93 (24)	24	100% schizophrenia	Inpatients	Translated	Greece	Reliability and validity of CDSS
Schuetze et al.	2001	0.76	0.93 (20)		20	Mixed psychotic disorders	Inpatients	Translated	Denmark	Reliability and validity of CDSS
Maggini & Raballo	2004		0.83 (24)		84	100% schizophrenia	Outpatients	Unknown	Italy	Measure of depression
Sarro et al.	2004	0.83	0.97 (27)	0.76 (14)	93	100% schizophrenia	Inpatients & Outpatients	Translated	Spain	Validation of CDSS
Kim et al.	2006	0.86			84	100% schizophrenia	Inpatients	Translated	Republic of Korea	Diagnostic validity of CDSS
Müller et al.	2006		≥0.87 (20)		119	100% schizophrenia	Inpatients	Translated	Germany	Sensitivity and specificity of CDSS
Zisook et al.	2006		≥0.90		165	Mixed psychotic disorders	Outpatients	Original	USA	Measure of depression
Coulston et al.	2007	0.81	0.87		59	Mixed psychotic disorders	Outpatients	Original	Australia	Measure of depression
Fitzgerald et al.	2008		>0.90 (20)		20	Mixed psychotic disorders	Outpatients	Original	Australia	Measure of depression
Jager et al.	2008		>0.80		288	100% schizophrenia	Inpatients	Translated	Germany	Measure of depression
Liu et al.	2009	0.8			101	100% schizophrenia	Inpatients	Translated	China	Diagnostic validity of CDSS
Xiao et al.	2009	0.8	0.88 (26)	0.93 (26)	26	100% schizophrenia	Inpatients	Translated	China	Reliability and validity of CDSS
Lincoln et al.	2010		0.92		80	Mixed psychotic disorders	Inpatients & Outpatients	Translated	Germany	Measure of depression
Chengappa et al.	2012		≥0.80		70	Mixed psychotic disorders	Outpatients	Original	USA	Measure of depression
Peleikis et al.	2013	0.82			128	Mixed psychotic disorders	Inpatients & Outpatients	Unknown	Norway	Measure of depression
Schennach et al.	2012	0.78	>0.80		278	Mixed psychotic disorders	Inpatients	Unknown	Germany	Comparison of CDSS with another measure
Moore et al.	2013	0.83			72	Mixed psychotic disorders	Outpatients	Original	USA	Measure of depression
Rabany et al.	2013	0.83			184	100% schizophrenia	Inpatients & Outpatients	Original	Israel	Measure of depression
Suttajit et al.	2013	0.87	0.98 (10)	0.86	60	100% schizophrenia	Inpatients & Outpatients	Translated	Thailand	Reliability and validity of CDSS
Ucok et al.	2013		0.83 (20)		103	100% schizophrenia	Outpatients	Translated	Turkey	Measure of depression
Bull et al.	2016	0.82			148	Mixed psychotic disorders	Outpatients	Unknown	Norway	Measure of depression
Garcia et al.	2016	0.87			79	Mixed psychotic disorders	Outpatients	Unknown	Spain	Measure of depression
Hani et al.	2016	0.82	0.90 (21)	0.85 (19)	102	100% schizophrenia	Inpatients	Translated	Qatar	Validation of CDSS
Krupchanka & Katliar	2016	0.88			96	100% schizophrenia	Inpatients & Outpatients	Translated	Belarus	Measure of depression
Mingrone et al.	2016		0.85 (15)		147	Mixed psychotic disorders	Outpatients	Unknown	Italy	Measure of depression
Pawelczyk et al.	2016		>0.82		71	100% schizophrenia	Inpatients	Unknown	Poland	Measure of depression
Grover et al.	2017	0.88	0.83 (42)		267	100% schizophrenia	Inpatients & Outpatients	Translated	India	Factor analysis of CDSS

(*continued on next page*)

Table 2 (*continued*)

Study	Year	Internal consistency	Inter-rater reliability (N)	Test-retest reliability (N)	N	Sample diagnosis	Sample Type	Version of the CDSS	Country in which study conducted	Context of CDSS use
Jeon et al.	2018		>0.75		56	100% schizophrenia	Outpatients	Unknown	Republic of Korea	Measure of depression
Xu et al.	2018	0.83	>0.85		348	100% schizophrenia	Inpatients	Translated	China	Measure of depression
Fathian et al.	2019		0.92		208	Mixed psychotic disorders	Inpatients	Original	Norway	Measure of depression
Richter et al.	2019	0.79			55	Mixed psychotic disorders	Outpatients	Unknown	Germany	Reliability and validity of another measure
Rostami et al.	2019	0.86		0.82	95	100% schizophrenia	Inpatients & Outpatients	Persian	Iran	Reliability and validity of CDSS

[a] Not all studies provided values for all types of reliability, so only reported values are included in this table. Some studies did not report whether the inter-rater reliability coefficients were calculated on the whole sample or a subset, as such where this information was reported it has been included in this table.

appropriate to use as a measure of between studies variation.

Fig. 3 reports the random effects synthesis of these 27 studies, with an estimated internal consistency of alpha = 0.83 (95% CI:0.82–0.84). There was an acceptable level of agreement between the primary studies' alpha estimates (Higgins $I^2 = 11.00\%$).

It should be noted that all primary studies reported internal consistency coefficients in excess of $\alpha = 0.70$.

3.4.1. The impact of influential primary studies

The impact of disproportionately influence studies was assessed using a "leave-one-out" analysis, in which the random effects model was calculated with each of the primary studies removed in turn. Change in weighted average effect size (i.e., influence) and the change in heterogeneity (i.e., discrepancy) was recorded. Fig. 4 presents the result of this "leave-one-out" analysis.

Fig. 4 reveals two studies (Grover et al., 2017; Schennach et al., 2012) as influential and discrepant. The study conducted by Schennach et al. (2012) utilised a large sample size as part of a multi-centre programme, with the authors noting they had "liberal" (pp. 284) inclusion criteria. From the reported information it was unclear what proportion of the sample had a diagnosis of schizophrenia, rather than another psychotic disorder. This suggests there may be a substantial amount of variation within the sample, potentially contributing to findings discrepant from the main body of literature. Grover et al.'s (2017) study was the only study conducted in India and using the Hindu version of the CDSS. Research has highlighted cultural differences in how depression presents, with variation between Eastern and Western cultures (Raguram et al., 2001). It may be that the questions within the CDSS are less applicable to this population. Therefore, the overall synthesis was recalculated with these two studies removed. This resulted in a negligible difference to the overall synthesis ($\alpha = 0.83$; 95% CI:0.81–0.84), and their exclusion did not change the overall conclusion of the analysis.

3.4.2. Attenuation due to risk of bias

The Quality Effects Model (Doi and Thalib, 2008) was calculated using the risk of bias ratings (supplementary Table 1) as weightings for methodological quality. This model reported an effect $\alpha = 0.83$ (95% CI:0.82–0.84), with heterogeneity remaining acceptable ($I^2 = 10.85\%$). This indicates the ratings of methodological quality of the studies does not result in a substantive change in the conclusion. Therefore, the conclusions of this meta-analysis can be considered robust to the ratings of methodological quality of the primary studies.

3.4.3. Publication bias and small study effects

Publication bias is caused by the tendency for statistically significant results to be published and to not publish papers with non-significant findings. Small study bias is the tendency for studies with smaller sample sizes to show greater variability in their measurement of internal consistency. These biases can be identified a funnel plot which plots the magnitude of the study's alpha estimate against the square root of the study's precision (sampling variances $= \frac{\sigma}{\sqrt{N}}$). If there is an absence of publication bias, the effects from the studies with small sample sizes which show greater variability will scatter more widely at the bottom of the plot compared to studies with larger samples at the top which will lie closer to the overall meta-analytic effect, creating a symmetrical funnel shape. If there is an absence of studies in the area of the plot associated with small sample sizes and non-significant results (for this meta-analysis it will be the bottom right-hand corner) then it is likely there is some publication bias leading to an overestimation of the true effect. The funnel plot of the z transformed alpha coefficients is presented in Fig. 5.

Fig. 5 does not suggest any obvious evidence of publication bias given the individual data points are symmetrical around the meta-analytic effect value. Additionally, Orwin's (Orwin, 1983) method indicates 131 studies reporting alpha coefficients of $\alpha < 0.6$ (i.e. 327.5% of the existing literature) would be required to reduce this effect below an alpha value of 0.7. These data can, therefore, be considered as robust to the effects of publication bias.

3.4.4. Attenuation due to diagnosis

To explore whether there was any influence of the sample's diagnosis on the distribution of alpha coefficients resulting from rating of the CDSS items, a subgroup analysis was conducted. This analysis grouped the primary studies into two groups: Pure and Mixed. 'Pure' refers to primary studies in which the entire sample had a reported diagnosis of paranoid schizophrenia. 'Mixed' refers to studies involving samples with a mixture of psychotic disorder diagnoses, including schizophrenia and schizoaffective disorder. The results of this analysis are presented in Table 4. The non-significant difference reflects a lack of a substantive difference in the estimates of reliability between both groups, such that it is unlikely that meaningful differences exist.

3.4.5. Attenuation due to version of measure

This subgroup analysis explored the influence of the language version of the CDSS that was used. This analysis grouped the primary studies into two groups: Original and Translated. 'Original' includes primary studies that utilised the original version of the CDSS (Addington et al., 1990). 'Translated' refers to all primary studies utilising a version of the CDSS that has been translated into another language. The results of this analysis are presented in Table 5. The non-significant difference reflects consistency in the effect across different language versions of the CDSS.

3.5. Inter-rater reliability

There were 28 studies reporting IRR coefficients. These were reported as kappa or ICC values, except for two papers as explained previously (Section 2.3 Data Extraction). For the purposes of analysis, IRR

Table 3
Quality framework for assessing risk of bias.

Area	Brief description	Risk of bias
Selection bias	The study sample is representative of that for which the CDSS was designed. The CDSS is specifically designed for adults with a diagnosis of Schizophrenia. It is suitable for individuals in both the acute and residual stages of the illness.	**Low risk:** The characteristics of the study population are clearly described and are representative of the population for which the scale was developed. 100% of participants have a diagnosis of Schizophrenia. **Unclear risk:** The characteristics of the study population are not clearly reported so it is unclear the proportion of the sample with a diagnosis of Schizophrenia, or the percentage of the sample with a diagnosis of Schizophrenia is less than 100% but greater than 50%. **High risk:** The sample characteristics are not representative for the scale's target population with less than 50% having a diagnosis of Schizophrenia.
Performance bias	Takes into consideration any alterations made to the original measure and the use of the scale. The following is outlined for use of the CDSS: the rater should have experience with individuals with Schizophrenia; it is to be administered as an interview with the 9th question based on rater observation; IRR with another rater experienced in using structured assessments should be developed; adequate IRR should be established within 5–10 practice interviews.	**Low risk:** The full version of the scale is used, either the original version or a version approved by the scale's developer (e.g. language variant). The measure was used in accordance with the scale protocol around training and administration. **Unclear risk:** It is unclear whether the full scale was administered, or it is unclear whether it is an approved version; or it is unclear whether the administration protocol was followed. **High risk:** Only selected items of the scale were administered; the scale's developer had not approved the version; or the administration protocol was not adhered to.
Reporting bias	Captures the completeness of the reporting within the study, around descriptive statistics and outcomes.	**Low risk:** There is a complete account of the descriptive statistics, with all results reported in full and appropriately. **Unclear risk:** Descriptive statistics are reported but are only partially reported. **High risk:** There are either no descriptive statistics or important data is missing within the reported dataset (e.g. data they said they were going to report has not been included).
Detection bias	Consideration of the detection of depression, as guided by the reported IRR. For this purpose, the acceptability of IRR coefficient values was determined using Koo and Li's (2016) guidelines.	**Low risk:** IRR coefficient was reported and is an acceptable value (>0.75). **Unclear risk:** IRR coefficient is not reported. **High risk:** IRR coefficient is reported and falls below the level considered acceptable (i.e. <0.75).
Statistical bias	The reporting of statistical information, relating to the reliability coefficient.	**Low risk:** Exact reliability coefficient is reported, and it is clear how this was calculated (i.e. no missing data). **Unclear risk:** Non-exact reliability coefficient is reported; or some data is missing (i.e. unclear whether the full sample was used to provide this value or just a subset of the sample). **High risk:** No information is provided as to how the reliability coefficient has been calculated.
Generalisability	Capturing the size of the sample and the ability to transfer findings to the wider population.	**Low risk:** The sample contains more than 40 participants. **Unclear risk:** The sample contains between 20 and 40 participants. **High risk:** The sample contains fewer than 20 participants.

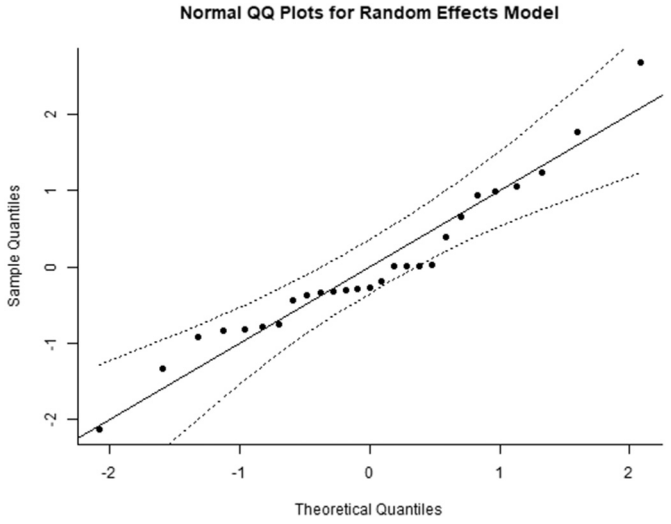

Fig. 2. QQ plot indicating normal distribution of study level effects for internal reliability coefficients.

values were in the form of kappa or ICC values with these being treated equivocally. The data arises from a total sample size of 1976 participants.[2] Sample sizes in the primary studies ranged from 10, accounting for five of the studies, to 349 (Xu et al., 2018). The distribution of the study level estimates of Fisher's (Fisher, 1921) transformation of IRR coefficients conformed to normal expectations, as shown in the QQ plot in Fig. 6. Given the normal distribution, the DerSimonian-Laird (DerSimonian and Laird, 1986) estimate is appropriate to use as the measure of between-studies variation.

The random effects synthesis of these 28 studies reported an estimated IRR of 0.88 (95% CI: 0.86–0.91), shown in Fig. 7. Although a high level of heterogeneity was observed (Higgins $I^2 = 78\%$, $tau^2 = 0.06$, $p < 0.01$), it should be noted that all of the primary studies reported a IRR coefficient greater than 0.7.

[2] It is of note that for 11 of the primary studies it was unclear whether the whole sample or a subset of the entire sample was used to calculate the inter-rater reliability value. For those studies for which it was unclear, the total sample has been used for the value of N.

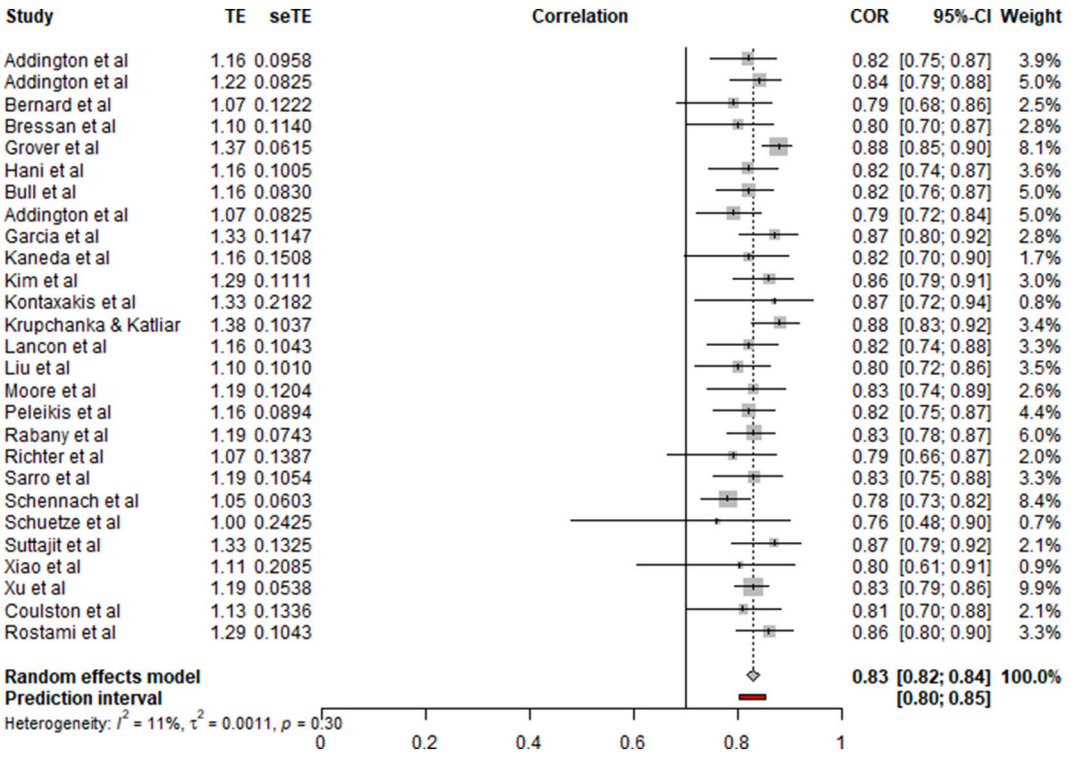

Fig. 3. Forest plot of the omnibus test of the internal reliability coefficients.

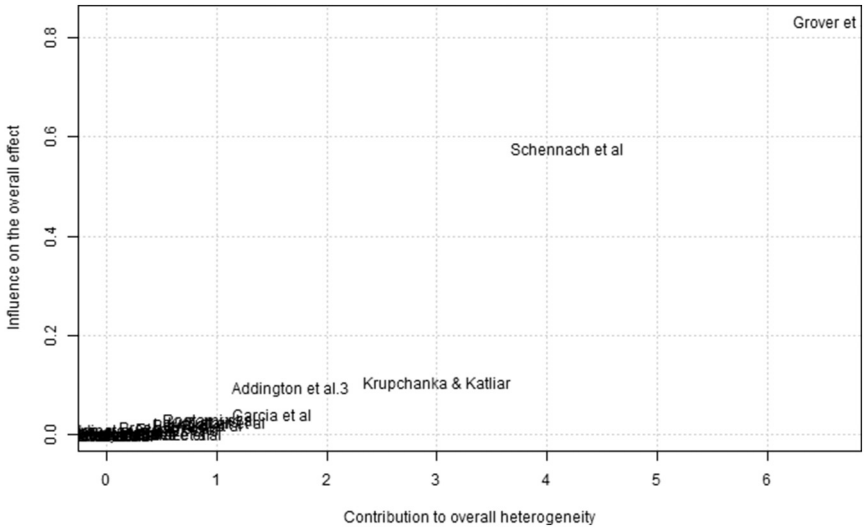

Fig. 4. Baujat (Baujat et al., 2002) scatterplot representing heterogeneity in estimates of internal consistency.

3.5.1. *Attenuation due to influential studies*

The Baujat (Baujat et al., 2002) scatterplot in Fig. 8 indicates there is one study (Bernard et al., 1998) that is very influential on the overall synthesis and is discrepant from the rest of the literature. This was the only study to report upon a French language version of the CDSS and the authors noted some participants had very low scores on the CDSS (mean: 6.97, range: 0–22).

When the meta-analysis was estimated with the Bernard et al. (1998) study removed, there was a slight decrease in the overall synthesis (estimated IRR = 0.87; 95% CI: 0.85–0.90) This, however, did not make any substantive difference to the overall conclusion of the analysis. As such, this study was retained in the dataset for calculating the overall effect.

3.5.2. *The impact of risk of bias in the primary studies*

The QEM was calculated using the risk of bias ratings (supplementary Table 1). The effect estimated by this model (QEM = 0.89, 95% CI: 0.87–0.92) suggests that studies with less risk of bias report slightly higher estimates of the IRR. Also, the estimate from the QEM is comparable to that of the non-weighted REM estimate, with only a negligible difference due to variation in risk of bias. These data are, therefore, considered robust to the effects of methodological bias.

3.5.3. *Publication bias and small study effects*

The association between the size of Fisher's (1921) z transformed correlation coefficient and the precision of measurement for each of the primary studies is reported in Fig. 9.

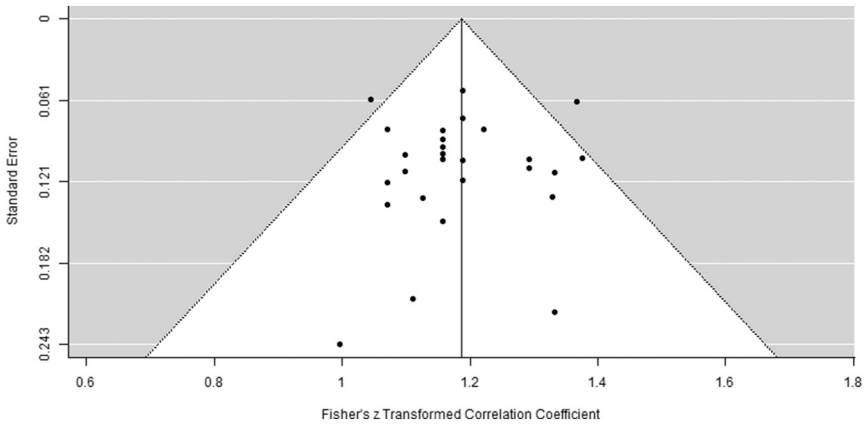

Fig. 5. Funnel plot of the z transformed alpha coefficients. The 95% confidence interval of the expected distribution of alpha is shown as an inverted "funnel."

Table 4
Subgroup analysis by composition of sample diagnosis.

Sub-group analysis		Number of studies	Random effects model		Heterogeneity			Between groups comparison
			Reliability	95% CI	Higgins I^2	Tau^2	Cochran's Q	
Nature of diagnosis of the study sample	Pure	19	0.84	0.82–0.85	6.20	0.0006	19.18	Q = 3.97,
	Mixed	8	0.81	0.79–0.83	0.00	0.00	6.01	p = 0.046

Table 5
Subgroup analysis by version of CDSS used.

Sub-group analysis		Number of studies	Random effects model		Heterogeneity			Between groups comparison
			Reliability	95% CI	Higgins I^2	Tau^2	Cochran's Q	
Version of CDSS used	Original	6	0.82	0.80–0.84	0.00	0.00	1.96	Q = 2.01,
	Translated	16	0.84	0.82–0.86	8.3	0.001	16.36	$p = 0.157$

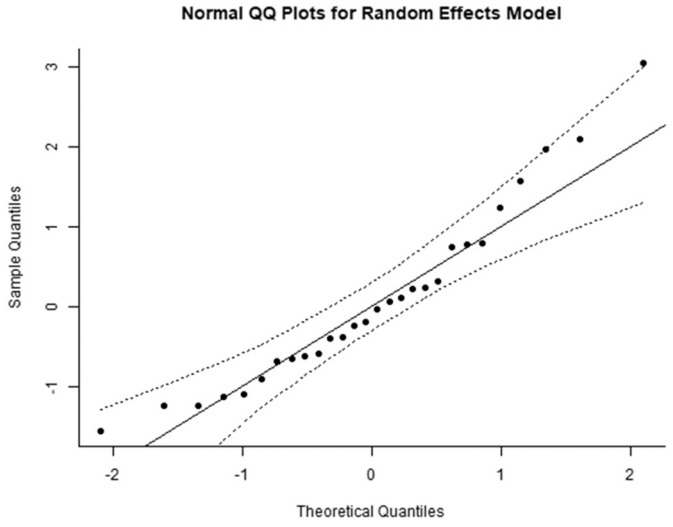

Fig. 6. QQ plot indicating normal distribution of study level effects for IRR coefficients.

The distribution of the data about the meta-analytic average suggests publication bias may be present. This is due to the lack of primary studies within the bottom left-hand corner of the funnel. It is of note, however, that using a trim and fill procedure (Duval and Tweedie, 2000a, 2000b) did not result in any corrections for publication bias. Some authorities have highlighted the potential limitations of relying on funnel plots alone for determining publication bias. This includes difficulties in correctly identifying publication bias from visual inspection of the funnel plot alone (Terrin et al., 2005). As such, Orwin's (1983) method was used alongside the funnel plot. This indicated 195 studies (i.e. 696% of the existing literature) reporting effect less than 0.6 would be required to reduce the effect below an IRR value of 0.70. These data can, therefore, be considered robust to the effects of publication bias.

3.5.4. *The impact of diagnosis*

To consider whether the composition of the sample's diagnosis had any influence on the distribution of IRR coefficients for the CDSS ratings, a subgroup analysis was conducted. The primary studies were categorised into two groups: Pure and Mixed, as occurred with the internal consistency data. The results of this analysis can be seen in Table 6. A non-significant and negligible difference was observed between the groups.

3.5.5. *Impact of the language of the CDSS*

The primary studies were categorised into two groups: Original and Translated, as occurred with the internal consistency data. Table 7 summarises the results of this analysis, which indicates there was a non-significant difference in the effect when analysed by the version of the CDSS used. As with the previous sub-group analysis (Table 6), this indicates consistency in the effect across the groups.

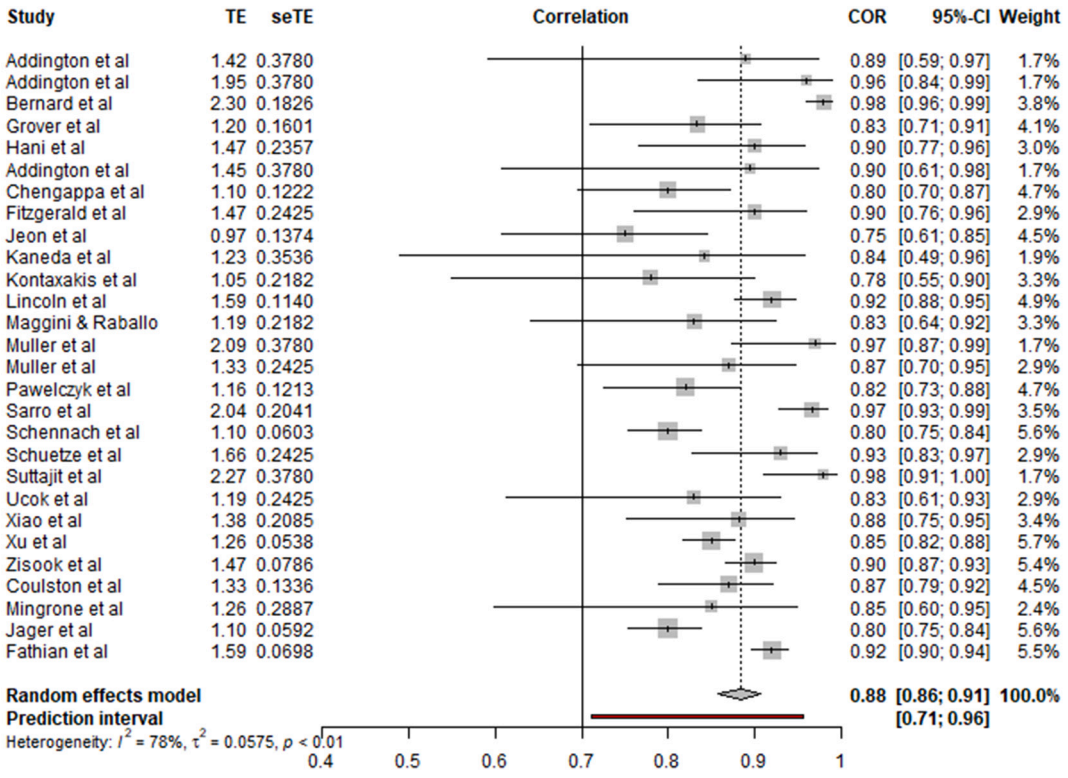

Fig. 7. Forest plot of the IRR coefficients.

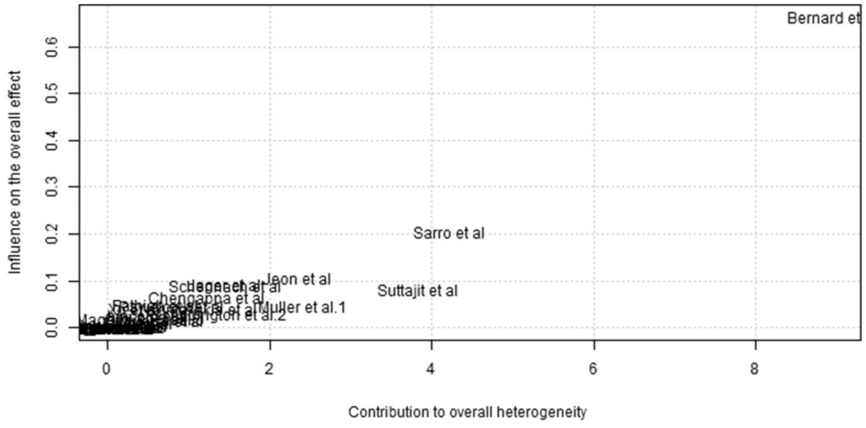

Fig. 8. Baujat scatterplot depicting heterogeneity in estimates of IRR.

3.6. Test-retest reliability

There were eight studies reporting test-retest reliability coefficients. These were reported as intraclass correlation values. The data arises from a total sample of 282,[3] with sample sizes ranging from 14 (Sarró et al., 2004) to 95 (Rostami et al., 2019). All studies had a maximum test-retest period of three days, apart from one (Kaneda et al., 2000) whose maximum test-retest period was eight days. The data regarding the test-retest period for Rostami et al.'s study was not reported.

The distribution of the study level estimates of Fisher's (1921) transformation of the test-retest reliability coefficients is shown in the QQ plot in Fig. 10. Despite the small number of studies, the data appears to approximate a normal distribution, such that the DerSimonian-Laird estimate is appropriate to use as the measure of between-studies variation.

Using this estimate, the random effects synthesis of these seven studies had an estimated test-retest reliability of 0.86 (95% CI:0.81–0.90), as shown in Fig. 11. The limited number of studies reporting test-retest reliability coefficients means that's no further statistical exploration of this effect can be undertaken.

4. Discussion

This meta-analytic review aimed to quantify the reliability of the CDSS. Of the studies identified as using the CDSS, 469 were excluded due to not reporting reliability data. A limited number of studies reported on the CDSS' test-retest reliability within their study (N = 8).

[3] It is of note that for three of the primary studies it was unclear whether the whole sample or a subset of the entire sample was used to calculate the test-retest reliability coefficient. For those studies for which it was unclear, the total sample has been used as the N value.

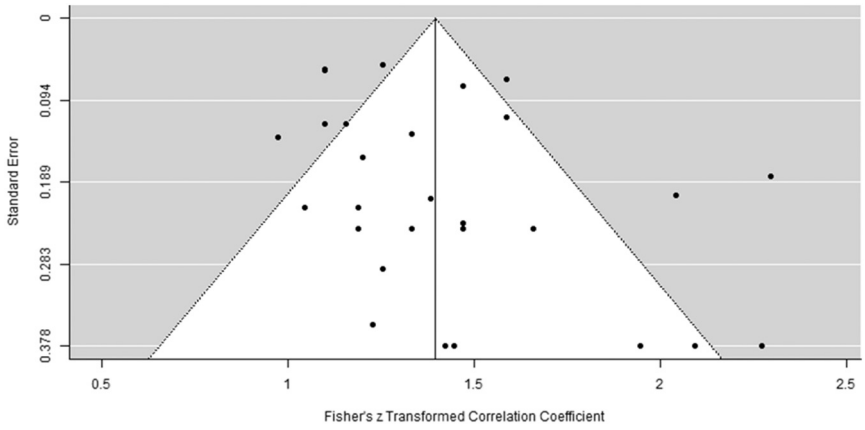

Fig. 9. Funnel plot of the correlation between Fisher's z transformed correlation coefficient and standard error (IRR coefficients).

Table 6
Subgroup analysis regarding sample diagnosis.

Sub-group analysis		Number of studies	Random effects model		Heterogeneity			Between groups comparison
			Reliability	95% CI	Higgins I^2	Tau^2	Cochran's Q	
Nature of diagnosis of the study sample	Pure	18	0.88	0.84–0.91	76.70	0.07	72.98	Q = 0.05,
	Mixed	10	0.89	0.87–0.92	80.40	0.05	45.84	$p = 0.821$

Table 7
Subgroup analysis for variant of the CDSS conducted.

Sub-group analysis		Number of studies	Random effects model		Heterogeneity			Between groups comparison
			Reliability	95% CI	Higgins I^2	Tau^2	Cochran's Q	
Version of CDSS used	Original	8	0.89	0.85–0.92	52.60	0.02	14.76	Q = 0.34,
	Translated	15	0.91	0.87–0.93	82.00	0.09	77.61	$p = 0.558$

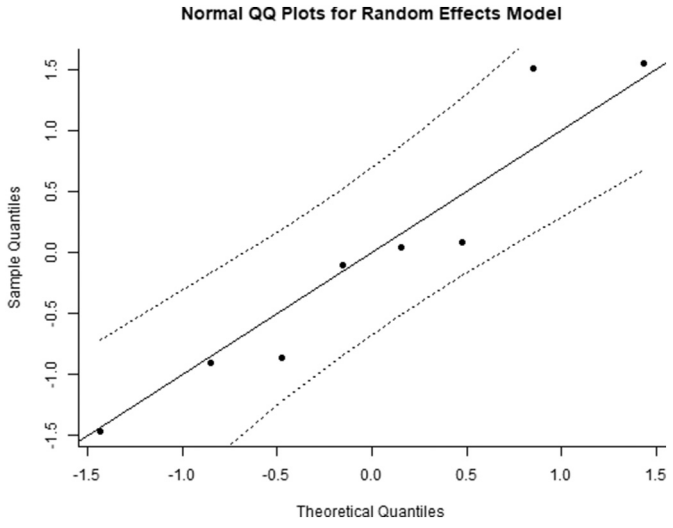

Fig. 10. QQ plot indicating distribution of study level effects for test-retest reliability coefficients.

Given the small number of studies, alongside the limitations of this literature (e.g. presence of heterogeneity and three studies having a small sample size (<20)), an in-depth analysis of test-retest reliability coefficients was not computed.

The review, therefore, focused on the CDSS' internal consistency and IRR. The search process identified 40 eligible studies, with 27 included within the meta-analysis of alpha coefficients and 28 used to meta-analyse IRR coefficients. The alpha coefficient meta-analysis (0.83, CI:0.82–0.84) suggests the CDSS has a good degree of internal consistency, in accordance with guidelines (Streiner, 2003) for interpreting alpha coefficients. Meta-analysing the ICC and kappa coefficients indicates the IRR of the CDSS is excellent (0.88; CI:0.86–0.91), when interpreted according to guidelines (Cichetti, 1994).

The current review highlighted non-significant differences in estimates of internal consistency (Q = 2.01, p = 0.157) and IRR (Q = 0.34, p = 0.58) between the original version of the CDSS and translated versions. Despite the CDSS being developed for individuals diagnosed with schizophrenia, rather than other psychotic disorders, this meta-analysis found consistency in the estimates of internal consistency and IRR between differently comprised samples (i.e. 100% vs. >50% had a diagnosis of schizophrenia). This may suggest the CDSS is detecting common aspects of symptoms across diagnoses.

A previous review (Lako et al., 2012) concluded that the CDSS had better reliability and validity than five other measures of depression. It reported the CDSS has good internal consistency (0.82; CI:0.76–0.88) and good IRR (0.86; CI: 0.73–0.98). These values are comparable to those of the current meta-analysis, albeit with larger CI, potentially reflecting the smaller number of studies upon which the estimates of Lako et al.'s (2012) review were based (N = 14). Lako et al.'s review (Lako et al., 2012) calculated reliability estimates using the mean of individual study coefficients, rather than weighted means as used in the current review. The simple arithmetic mean is subject to multiple biases, principle among which is the failure of the arithmetic mean to consider the precision of the individual estimates. This means poorer quality evidence is weighted as highly as better-quality evidence. The current review can, therefore, be considered the first of its kind to quantitatively synthesise the CDSS' internal consistency and IRR. This review also

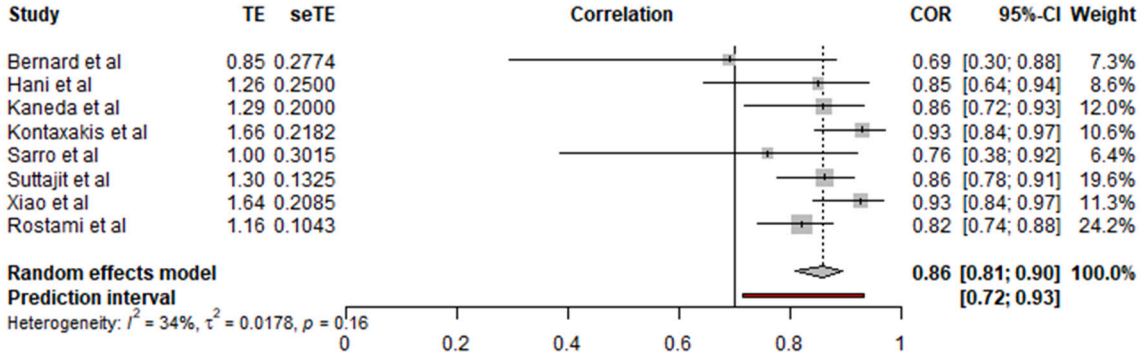

Fig. 11. Forest plot of the omnibus test of the test-retest reliability coefficients.

highlighted the CDSS to have good internal consistency and excellent IRR for multiple languages.

The findings present the CDSS as a psychometrically-sound tool, capable of screening individuals with schizophrenia for depressive symptoms. This review highlighted language variants of the CDSS as reliable with respect to internal consistency and IRR. This may enable patients to be assessed using a measure in their first language or for bilingual patients to select the language version with which they are more confident.

For both internal consistency and IRR, the meta-analytic effect is greater than the threshold recommended (i.e. >0.70 and >0.60 respectively). This means clinicians can be confident in using the CDSS to screen for symptoms of depression in individuals diagnosed with schizophrenia, and subsequently acting upon the findings. Given the findings regarding the translated versions, it may be beneficial for clinicians to consider which version may be best (i.e. more reliable) when working with bilingual patients.

For clinicians, it is important to consider whether an individual's change in score on a measure reflects a clinically significant difference. The Reliable Change Index provides a means of determining the degree of change required for an individual's presentation to be considered a reliable change in their mental wellbeing (Ferguson et al., 2002). Table 8 reports the cut-offs for reliable change for inpatient and outpatient respondents using standard deviations reported by Addington et al. (1994) and the current review's alpha coefficients. Reliable change was calculated using procedures described by Jacobson and Truax (1991). This indicates for an alpha coefficient of 0.83, a change of 5.14 (inpatients) or 4.64 (outpatients) to an individual's CDSS score is required for clinicians to be sufficiently confident (i.e. within 95% CI) that the individual is showing a clinically reliable change in their experience of depressive symptoms.

This meta-analysis focused upon the CDSS' internal consistency and IRR. Another area of importance is the tool's stability across time. As mentioned, eight studies were identified that reported test-retest coefficients. The quality of the data alongside the small body of literature was considered unsuitable for a meta-analytic review (Rostami et al., 2019).

One of the exclusion criteria involved filtering out primary studies in a non-English language. This resulted in 31 studies being excluded, although it was unclear whether these papers provided the required data for inclusion. As translated versions of the CDSS may be published in non-English journals, there is potentially a number of relevant articles not included within this review. A future review would benefit from including non-English studies to provide a more comprehensive insight into the reliability of translated versions of the CDSS.

In conclusion, this review quantified the CDSS' internal consistency and the IRR using advanced statistical techniques. The findings suggest it is a reliable tool for assessing symptoms of depression in individuals with schizophrenia. This provides reassurance to clinicians and patients that it is an appropriate measure to use within clinical practice to distinguish between symptoms of depression and schizophrenia, and, therefore, inform treatment efforts.

Table 8
Reliable change indices at varying levels of confidence.

	Inpatient	Outpatient
Reliable change at 51% confidence	±1.73	±1.56
Reliable change at 66% confidence	±2.62	±2.37
Reliable change at 95% confidence	±5.14	±4.64
Reliable change at 99% confidence	±5.25	±4.73

This review also suggested comparable reliability estimates across the language variants of the CDSS. This was via a subgroup analysis of studies using the original measure against those using a translated version. To develop understanding of the psychometric properties of these versions, further research may be beneficial as several translated versions were only used in one primary study. This may include consideration of whether there are different norms for different populations (e.g. age or diagnosis).

Future reviews of this nature would benefit from authors reporting reliability coefficients for assessment tools used within their study, regardless of the study's aims. This would expand the range of literature incorporated within such reviews. The ability to synthesise results from individual researchers would also be improved by clearly documenting how reliability coefficients were calculated. For example, data extraction in the current review was compromised by various studies not detailing how IRR had been computed (i.e. on what proportion of the sample and the number of raters). Improved data reporting will provide further opportunities to quantify the internal consistency and IRR of the CDSS and enable its test-retest reliability to be quantified.

Funding statement

This review was completed as part of a Doctor of Clinical Psychology and no funding was received from any organisations.

CRediT authorship contribution statement

Laura Porter, University of Birmingham.
Christopher Jones, University of Birmingham.
Andrew Fox, University of Birmingham.

Declaration of competing interest

The authors have declared that there are no conflicts of interest in relation to the subject of this article.

Acknowledgement

This meta-analysis was completed as part of a Doctorate of Clinical

Psychology with support from the University of Birmingham and Birmingham and Solihull Mental Health Foundation NHS Trust.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.schres.2021.11.040.

References

Addington, D., Addington, J., Schissel, B., 1990. A depression rating scale for schizophrenics. Schizophr. Res. 3 (4), 247–251. https://doi.org/10.1016/0920-9964(90)90005-R.

Addington, D., Addington, J., Maticka-Tyndale, E., Joyce, J., 1992. Reliability and validity of a depression rating scale for schizophrenics. Schizophr. Res. 6 (3), 201–208. https://doi.org/10.1016/0920-9964(92)90003-N.

Addington, D., Addington, J., Maticka-Tyndale, E., 1993. Assessing depression in schizophrenia: the Calgary depression scale. Br. J. Psychiatry 163 (S22), 39–44. https://doi.org/10.1192/S0007125000292581.

Addington, D., Addington, J., Maticka-Tyndale, E., 1994. Specificity of the Calgary depression scale for schizophrenics. Schizophr. Res. 11 (3), 239–244. https://doi.org/10.1016/0920-9964(94)90017-5.

Addington, D., Addington, J., Atkinson, M., 1996. A psychometric comparison of the Calgary depression scale for schizophrenia and the Hamilton depression rating scale. Schizophr. Res. 19 (2–3), 205–212. https://doi.org/10.1016/0920-9964(95)00070-4.

Baujat, B., Mahé, C., Pignon, J.-P., Hill, C., 2002. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials: GRAPHICAL METHOD FOR EXPLORING HETEROGENEITY IN META-ANALYSES. Stat. Med. 21 (18), 2641–2652. https://doi.org/10.1002/sim.1221.

Bernard, D., Lançon, C., Auquier, P., Reine, G., Addington, D., 1998. Calgary depression scale for schizophrenia: a study of the validity of a french-language version in a population of schizophrenic patients. Acta Psychiatr. Scand. 97 (1), 36–41. https://doi.org/10.1111/j.1600-0447.1998.tb09960.x.

Birchwood, M., Iqbal, Z., Chadwick, P., Trower, P., 2000. Cognitive approach to depression and suicidal thinking in psychosis. I. Ontogeny of post-psychotic depression. Br. J. Psychiatry 177, 516–528.

Bressan, R.A., Chaves, A.C., Shirakawa, I., Mari, J.de J., 1998. Validity study of the Brazilian version of the Calgary Depression Scale for Schizophrenia. Schizophr. Res. 32 (1), 31–39. https://doi.org/10.1016/S0920-9964(98)00029-2.

British Psychological Society (BPS), 2017. Understanding Psychosis and Schizophrenia. British Psychological Society, Leicester.

Buckley, P.F., Miller, B.J., Lehrer, D.S., Castle, D.J., 2009. Psychiatric comorbidities and schizophrenia. Schizophr. Bull. 35 (2), 383–402. https://doi.org/10.1093/schbul/sbn135.

Bull, H., Ueland, T., Lystad, J.U., Evensen, S., Martinsen, E.W., Falkum, E., 2016. Vocational functioning in schizophrenia Spectrum disorders: does apathy Matter? J. Nerv. Ment. Dis. 204 (8), 599–605. https://doi.org/10.1097/NMD.0000000000000504.

Chengappa, K.N., Turkin, S.R., DeSanti, S., Bowie, C.R., Brar, J.S., Schlicht, P.J., Fleet, D., 2012. A preliminary, randomized, double-blind, placebo-controlled trial of L-carnosine to improve cognition in schizophrenia. Schizophr. Res. 142, 145–152.

Cichetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assess. 6 (4), 284–290.

Conley, R., Aschersvanum, H., Zhu, B., Faries, D., Kinon, B., 2007. The burden of depressive symptoms in the long-term treatment of patients with schizophrenia. Schizophr. Res. 90 (1–3), 186–197. https://doi.org/10.1016/j.schres.2006.09.027.

Corey, D.M., Dunlap, W.P., Burke, M.J., 1998. Averaging correlations: expected values and bias in combined Pearson r s and Fisher's z transformations. J. Gen. Psychol. 125 (3), 245–261. https://doi.org/10.1080/00221309809595548.

Coulston, C., Perdices, M., Tennant, C., 2007. The neuropsychological correlates of cannabis use in schizophrenia: lifetime abuse/dependence, frequency of use, and recency of use. Schizophr. Res. 96 (1–3), 169–184. https://doi.org/10.1016/j.schres.2007.08.006.

Cuesta, M.J., Basterra, V., Sanchez-Torres, A., Peralta, V., 2009. Controversies surrounding the diagnosis of schizophrenia and other psychoses. Expert. Rev. Neurother. 9 (10), 1475–1486. https://doi.org/10.1586/ern.09.102.

DerSimonian, R., Laird, N., 1986. Meta-analysis in clinical trials. Control. Clin. Trials 7 (3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2.

Doi, S.A.R., Thalib, L., 2008. A quality-effects model for meta-analysis. Epidemiology 19 (1), 94–100. https://doi.org/10.1097/EDE.0b013e31815c24e7.

Duval, S., Tweedie, R., 2000a. A nonparametric, "Trim and fill" method of accounting for publication bias in meta-analysis. J. Am. Stat. Assoc. 95 (449), 89–98. https://doi.org/10.1080/01621459.2000.10473905.

Duval, S., Tweedie, R., 2000b. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56 (2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x.

Edwards, C.J., Garety, P., Hardy, A., 2019. The relationship between depressive symptoms and negative symptoms in people with non-affective psychosis: a meta-analysis. Psychol. Med. 49 (15), 2486–2498. https://doi.org/10.1017/S0033291719002381.

Fathian, F., Løberg, E.-M., Gjestad, R., Steen, V.M., Kroken, R.A., Jørgensen, H.A., Johnsen, E., 2019. Associations between C-reactive protein levels and cognition during the first 6 months after acute psychosis. Acta Neuropsychiatrica 31 (1), 36–45. https://doi.org/10.1017/neu.2018.25.

Ferguson, R.J., Robinson, A.B., Splaine, M., 2002. Use of the reliable change index to evaluate clinical significance in SF-36 outcomes. Qual. Life Res. 11, 509–516.

Fisher, R.A., 1921. On the 'probable' error' of a coefficient of correlation reduced from a small sample. Metron 1, 3–32.

Fitzgerald, P.B., Herring, S., Hoy, K., McQueen, S., Segrave, R., Kulkarni, J., Daskalakis, Z.J., 2008. A study of the effectiveness of bilateral transcranial magnetic stimulation in the treatment of the negative symptoms of schizophrenia. Brain Stimul. 1 (1), 27–32. https://doi.org/10.1016/j.brs.2007.08.001.

Foussias, G., Remington, G., 2010. Negative symptoms in schizophrenia: avolition and Occam's razor. Schizophr. Bull. 36 (2), 359–369. https://doi.org/10.1093/schbul/sbn094.

Garcia, M., Motalvo, I., Creus, M., Cabezas, Á., Solé, M., Algora, M.J., Labad, J., 2016. Sex differences in the effect of childhood trauma on the clinical expression of early psychosis. Compr. Psychiatry 68, 86–96.

Glen, S., 2014. "Cohen's Kappa Statistic" From StatisticsHowTo.com: Elementary Statistics for the rest of us!. Accessed March 20 from: https://www.statisticshowto.datasciencecentral.com/cohens-kappa-statistic/.

Grover, S., Sahoo, S., Dua, D., Chakrabarti, S., Avasthi, A., 2017. Scales for assessment of depression in schizophrenia: factor analysis of Calgary depression rating scale and Hamilton depression rating scale. Psychiatry Res. 252, 333–339. https://doi.org/10.1016/j.psychres.2017.03.018.

Hani, Y., Ghuloum, S., Mahfoud, Z., Opler, M., Khan, A., Yehya, A., Al-Amin, H., 2016. Validation of the Arabic version of the calgary depression scale for schizophrenia. PLoS ONE 11 (9), e0162304.

Higgins, J.P., Altman, D.G., Gøtzsche, P.C., Jüni, P., Moher, D., Oxman, A.D., Sterne, J.A., 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br. Med. J. 343, 1–9. https://doi.org/10.1136/bmj.d5928.

Jacobson, N.S., Truax, P., 1991. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J. Consult. Clin. Psychol. 59 (1), 12–19.

Jäger, M., Messer, T., Laux, G., Pfeiffer, H., Naber, D., Schmidt, L.G., Möller, H.J., 2008. Standardized remission criteria in schizophrenia: descriptive validity and comparability with previously used outcome measures. Pharmacopsychiatry 41, 190–195.

Jeon, D., Jung, D., Kim, S., Shim, J., Moon, J., Seo, Y., Kim, Y., 2018. Adjunct transcranial direct current stimulation improves cognitive function in patients with schizophrenia: a double-blind 12-week study. Schizophr. Res. 197, 378–385.

Kaneda, Y., Akira, F., Tetsuro, O., 2000. Psychometric properties of the japanese version of the Calgary depression scale for schizophrenia. J. Nerv. Ment. Dis. 188 (4), 237–239.

Kim, S.-W., Kim, S.-J., Yoon, B.-H., Kim, J.-M., Shin, I.-S., Hwang, M.Y., Yoon, J.-S., 2006. Diagnostic validity of assessment scales for depression in patients with schizophrenia. Psychiatry Res. 144 (1), 57–63. https://doi.org/10.1016/j.psychres.2005.10.002.

Kontaxakis, V.P., Havaki-Kontaxaki, B.J., Margariti, M.M., Stamouli, S.S., Kollias, C.T., Angelopoulos, E.K., Christodoulou, G.N., 2000. The greek version of the Calgary depression scale for schizophrenia. Psychiatry Res. 94, 163–171.

Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 15 (2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012.

Krupchanka, D., Katliar, M., 2016. The role of insight in moderating the association between depressive symptoms in people with schizophrenia and stigma among their nearest relatives: a pilot study. Schizophr. Bull. 42 (3), 600–607. https://doi.org/10.1093/schbul/sbw024.

Lako, I.M., Bruggeman, R., Knegtering, H., Wiersma, D., Schoevers, R.A., Slooff, C.J., Taxis, K., 2012. A systematic review of instruments to measure depressive symptoms in patients with schizophrenia. J. Affect. Disord. 140, 38–47.

Lançon, C., Auquier, P., Reine, G., Toumi, M., Addington, D., 1999. Evaluation of depression in schizophrenia: psychometric properties of a french version of the Calgary depression scale. Psychiatry Res. 89 (2), 123–132. https://doi.org/10.1016/S0165-1781(99)00097-9.

Lincoln, T.M., Ziegler, M., Lüllmann, E., Müller, M.J., Rief, W., 2010. Can delusions be self-assessed? Concordance between self- and observer-rated delusions in schizophrenia. Psychiatry Res. 178 (2), 249–254. https://doi.org/10.1016/j.psychres.2009.04.019.

Liu, H., Zhang, H., Xiao, W., Liu, Q., Fu, P., Chen, J., Li, L., 2009. Scales for evaluating depressive symptoms in Chinese patients with schizophrenia. J. Nerv. Ment. Dis. 197 (2), 140–142.

Maggini, C., Raballo, A., 2004. Self-centrality, basic symptoms model and psychopathology in schizophrenia. Psychopathology 37 (2), 69–75. https://doi.org/10.1159/000077581.

Mingrone, C., Montemagni, C., Sandei, L., Bava, I., Mancini, I., Cardillo, S., Rocca, P., 2016. Coping strategies in schizoaffective disorder and schizophrenia: differences and similarities. Psychiatry Res. 244, 317–323.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Open Med. 3 (2), 123–130.

Moore, R.C., Viglione, D.J., Rosenfarb, I.S., Patterson, T.L., Mausbach, B.T., 2013. Rorschach measures of cognition relate to everyday and social functioning in schizophrenia. Psychol. Assess. 25 (1), 253–263. https://doi.org/10.1037/a0030546.

Müller, M.J., Marx-Dannigkeit, P., Schlösser, R., Wetzel, H., Addington, D., Benkert, O., 1999. The Calgary depression rating scale for schizophrenia: development and

interrater reliability of a german version (CDSS-G). J. Psychiatr. Res. 33 (5), 433–443. https://doi.org/10.1016/S0022-3956(99)00018-7.
Müller, M.J., Müller, K.-M., Fellgiebel, A., 2006. Detection of depression in acute schizophrenia: sensitivity and specificity of 2 standard observer rating scales. Can. J. Psychiatry 51 (6), 387–392. https://doi.org/10.1177/070674370605100609.
NICE, 2014. Psychosis and schizophrenia in adults: prevention and management. In: Clinical Guideline 178.
Orwin, R.G., 1983. A fail-safe N for effect size in meta-analysis. J. Educ. Stat. 8, 157–159.
Pawełczyk, T., Grancow-Grabka, M., Kotlicka-Antczak, M., Trafalska, E., Pawełczyk, A., 2016. A randomized controlled study of the efficacy of six-month supplementation with concentrated fish oil rich in omega-3 polyunsaturated fatty acids in first episode schizophrenia. J. Psychiatr. Res. 73, 34–44. https://doi.org/10.1016/j.jpsychires.2015.11.013.
Peleikis, D.E., Varga, M., Sundet, K., Lorentzen, S., Agartz, I., Andreassen, O.A., 2013. Schizophrenia patients with and without post-traumatic stress disorder (PTSD) have different mood symptom levels but same cognitive functioning. Acta Psychiatr. Scand. 127 (6), 455–463. https://doi.org/10.1111/acps.12041.
Price, P.C., Jhangiani, R., Chiang, I.A., 2015. Chapter 5: psychological measurement. In: Price, P.C., Jhangiani, R., Chang, I.A. (Eds.), Research Methods in Psychology. BC Campus, Victoria, B.C., pp. 80–99
Rabany, L., Weiser, M., Levkovitz, Y., 2013. Guilt and depression: two different factors in individuals with negative symptoms of schizophrenia. European Psychiatry 28 (6), 327–331. https://doi.org/10.1016/j.eurpsy.2012.02.008.
Raguram, R., Weiss, M.G., Keval, H., Channabasavanna, S.M., 2001. Cultural dimensions of clinical depression in Bangalore, India. Anthropol. Med. 8 (1), 31–46. https://doi.org/10.1080/13648470120063889.
Rector, N.A., Beck, A.T., Stolar, N., 2005. The negative symptoms of schizophrenia: a cognitive perspective. Can. J. Psychiatry 50 (5), 247–257. https://doi.org/10.1177/070674370505000503.
Richter, J., Hesse, K., Eberle, M., Eckstein, K.N., Zimmermann, L., Schreiber, L., Klingberg, S., 2019. Self-assessment of negative symptoms – critical appraisal of the motivation and pleasure – self-report's (MAP-SR) validity and reliability. Compr. Psychiatry 88, 22–28.
Rodriguez, M.C., Maeda, Y., 2006. Meta-analysis of coefficient alpha. Psychol. Methods 11 (3), 306–322. https://doi.org/10.1037/1082-989X.11.3.306.
Rostami, R., Kazemi, R., Khodaie-Ardakani, M., Sohrabi, L., Ghiasi, S., Kamali, Z.S., Addington, D.E., 2019. The Persian version of the Calgary Depression Scale for Schizophrenia (CDSS-P). Asian J. Psychiatr. 45, 44–49.
Sarró, S., Duenas, R., Ramírez, N., Arranz, B., Martínez, R., Sánchez, J.M., San, L., 2004. Cross-cultural adaptation and validation of the Spanish version of the Calgary Depression Rating Scale for Schizophrenia. Schizophr. Res. 68, 349–356.
Schennach, R., Obermeier, M., Seemüller, F., Jäger, M., Schmauss, M., Laux, G., Möller, H., 2012. Evaluating depressive symptoms in schizophrenia: a psychometric comparison of the Calgary depression scale for schizophrenia and the Hamilton depression rating scale. Psychopathology 45, 276–285.
Schuetze, T., Nørholm, V., Olsen, L.R., Hougaard, H., Ekstrøm, M., Wagn, P., Bech, P., 2001. Reliability and validity of the danish version of the Calgary depression scale for schizophrenia. Nord. J. Psychiatry 55, 119–122.
Smith, J., Noble, H., 2014. Bias in research: table 1. Evid. Based Nurs. 17 (4), 100–101. https://doi.org/10.1136/eb-2014-101946.
Streiner, D.L., 2003. Starting at the beginning: an introduction to coefficient alpha and internal consistency. J. Pers. Assess. 80 (1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18.
Suttajit, S., Srisurapanont, Pilakanta, Charnsil, Suttajit, 2013. Reliability and validity of the Thai version of the Calgary Depression Scale for Schizophrenia. Neuropsychiatr. Dis. Treat. 113. https://doi.org/10.2147/NDT.S40292.
Terrin, N., Schmid, C.H., Lau, J., 2005. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J. Clin. Epidemiol. 58 (9), 894–901. https://doi.org/10.1016/j.jclinepi.2005.01.006.
Üçok, A., Karadayı, G., Emiroğlu, B., Sartorius, N., 2013. Anticipated discrimination is related to symptom severity, functionality and quality of life in schizophrenia. Psychiatry Res. 209 (3), 333–339. https://doi.org/10.1016/j.psychres.2013.02.022.
Upthegrove, R., Marwaha, S., Birchwood, M., 2016. Depression and schizophrenia: cause, consequence or trans-diagnostic issue? Schizophr. Bull. sbw097. https://doi.org/10.1093/schbul/sbw097.
van Os, J., Gilvarry, C., Bale, R., van Horn, E., Tattan, T., White, I., Murray, R., 2000. Diagnostic value of the DSM and ICD categories of psychosis: an evidence-based approach. Soc. Psychiatry Psychiatr. Epidemiol. 35, 305–311.
Walker, E., Hernandez, A.V., Kattan, M.W., 2008. Meta-analysis: its strengths and limitations. Cleve. Clin. J. Med. 75 (6), 431–439. https://doi.org/10.3949/ccjm.75.6.431.
Xiao, W., Liu, H., Zhang, H., Liu, Q., Peixin, F., Chen, J., Shu, L., 2009. Reliability and validity of the Chinese version of the Calgary Depression Scale for Schizophrenia. Aust. N. Z. J. Psychiatry 43, 548–553.
Xu, Y., Li, F., Liu, X., Zhong, B., 2018. Depressive symptoms in chinese male inpatients with schizophrenia: prevalence and clinical correlates. Psychiatry Res. 264, 380–384.
Zisook, S., Nyer, M., Kasckow, J., Golshan, S., Lehman, D., Montross, L., 2006. Depressive symptom patterns in patients with chronic schizophrenia and subsyndromal depression. Schizophr. Res. 86 (1–3), 226–233. https://doi.org/10.1016/j.schres.2006.03.047.