

Psychometric Properties of WHO's Schedules for Clinical Assessment in Neuropsychiatry: A Systematic Review.

Wubalem Fekadu

wubalem.fekadu@aau.edu.et

Addis Ababa University <https://orcid.org/0000-0001-7418-4955>

Awoke Mihretu

Addis Ababa University

Atalay Alem

Addis Ababa University

Traolach Brugha

University of Leicester, Leicester, UK

Mark van Ommeren

World Health Organization

Somnath Chatterji

World Health Organization

Charlotte Hanlon

University of Edinburgh

Abebaw Fekadu

Addis Ababa University

Systematic Review

Keywords: Schedules for Clinical Assessment in Neuropsychiatry, Psychometric Properties, Systematic review

Posted Date: September 30th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7730040/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Background

WHO's Schedules for Clinical Assessment in Neuropsychiatry (SCAN) is often used as the gold standard for psychiatric classification. We systematically reviewed studies on the psychometric properties of the SCAN to support its adaptation to the revised international classification systems.

Methods

We searched PubMed, PsycINFO, Embase, Global Health, and Global Index Medicus up to April 17, 2025, and contacted experts. The protocol was registered in PROSPERO (CRD42024522395).

Results

Titles and abstracts of 4,241 records were screened, with 296 full-text articles evaluated. Ninety-three articles were included in the final review: 46 assessing SCAN's psychometric properties and 47 validating other measures using SCAN as a gold standard. The internal consistency of the SCAN and its predecessor, the Present State Examination (PSE), ranged from good to excellent. Both demonstrated acceptable intra-rater, inter-rater, and test-retest reliability, with reliability especially high for psychotic disorders. There was also evidence supporting concurrent, construct, semantic, and content validity, although there was an absence of evidence for predictive validity. We also found acceptable psychometric properties for the different syndrome-based sections of the SCAN.

Conclusion

Although recent, high-quality studies are scarce, the SCAN is a promising tool for diagnosing a variety of psychiatric issues, particularly psychotic disorders. It demonstrates established reliability and evidence of concurrent, construct, semantic, and content validity. However, there is a need to revise the current version of SCAN to align it with contemporary diagnostic systems. Additionally, further research is required, especially regarding the assessment of non-psychotic conditions.

1. Introduction

Structured mental disorders' classification instruments remain fundamental tools for mental health research and evidence-based practice. While simple self-reported symptom scales are commonly used to determine the occurrence of mental health symptoms and track the course of illness, classification tools are critical for accurately identifying illness conditions, delineating co-morbidity, and assisting in planning services and better care [1].

The Schedules for Clinical Assessment in Neuropsychiatry (SCAN) is a semi-structured World Health Organization (WHO) assessment interview tool that systematises the clinical examination, the elicitation of psychopathology to match a standard definition in a glossary, and the classification of neuropsychiatric conditions [2–4]. SCAN allows the identification of disorders according to standard international diagnostic systems, through accompanying algorithms that interpret the clinical descriptions or diagnostic criteria of these systems, and offers the possibility of dimensional categorization of psychopathology [2] .

The development of SCAN began in the late 1950s, with the Present State Examination, and has since been revised several times. These revisions have remained true to the fundamental commitment of the SCAN to ensuring accurate assessment, systematisation, and transcultural applicability [3–6]. It has been translated into more than 35 languages [2]

SCAN is a collection of instruments designed to assess, measure, and classify psychopathology. It consists of four main components: the Present State Examination, the Glossary of Differential Definitions, the Item Group Checklist, and the Clinical History Schedule. The Present State Examination is divided into two parts: Part One focuses on non-psychotic disorders, while Part Two addresses psychotic and cognitive disorders [2].

The SCAN offers flexibility in constructing a clinical database which is not confined to any specific nosology. This approach enables adaptability in the interview process, allowing for modifications in question order and wording based on the course of the interview. Interviewers have the freedom to explore certain lines of inquiry while bypassing others as needed. Importantly, the examiner determines the presence of symptoms rather than relying solely on the person's report.

SCAN has been used in efforts to examine the manifestations and epidemiology of psychosis in the Global South [7, 8] and for trials of interventions for people with psychosis in Africa [9]. It was also used in Indigenous Andean culture [10]. SCAN is used every seven years in England in the Adult Psychiatric Morbidity Survey (APMS) as a second-stage (confirmatory) examination in persons in the community when psychosis is suspected [11].

It is being utilized in a global study across 19 sites in 12 countries throughout Asia, Africa, the Americas, and Europe. The aim is to understand the molecular and systemic factors linking COVID-19 to both short-term and long-term neurological illnesses [12]. Although studies have evaluated the psychometric properties of SCAN [6, 13] or some of its sections [14], there are also questions about its design, the requirement for specialized training, and the perceived unfriendliness of the format [1]. Moreover, with the availability of new diagnostic instruments since the development of the PSE (and the SCAN), for example, the Structured Clinical Interview for DSM (SCID), it is important to revisit its ongoing relevance in a way that informs the SCAN revision for ICD-11 and the broader advancement of psychiatric research enabled by technology.

We conducted a systematic review to establish the psychometric properties of the various revisions of the SCAN, including its predecessor, the Present State Examination (PSE). We also profiled the studies that used the SCAN as a gold standard in validating other diagnostic and screening measures.

2. Methods

We adhered to the updated Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline to conduct and report this systematic review [15] and the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) criteria [16]. These guidelines cover reporting, data extraction, methodological quality criteria, and risk of bias assessment [17]. The protocol was registered on the PROSPERO International Register of Systematic Reviews (CRD42024522395).

2.1. Search strategy

We searched five databases: PubMed, Embase, PsycINFO, Global Health, and Global Index Medicus. The databases were searched from the inception of each database with no language restriction up to April 17, 2024. Forward and backward searches were conducted to identify additional relevant studies. We also 'snowballed based on recommendations of people who had published using SCAN. The search terms consisted of keywords, MeSH and Emtree terms, and vocabularies for SCAN and psychometric properties. The terms were combined with the Boolean term AND:

Terms

1. **Terms for SCAN:** SCAN OR SCAN 1•0 OR SCAN 2 OR SCAN 2•1 OR "Schedules for Clinical Assessment in Neuropsychiatry" OR "Schedules for Clinical Assessment in Neuropsychiatry 1.0" OR "Schedules for Clinical Assessment in Neuropsychiatry 2.0" OR "Schedules for Clinical Assessment in Neuropsychiatry 2.1" OR PSE OR "Present State Examination" OR "MINI SCAN"
2. **Terms for psychometric Properties:** Psychometric* OR Validity OR Reliability OR "Reproducibility of Results" (Supplementary file_1).

2.2. Eligibility criteria

We included studies that met the following criteria:

1. Validation studies that reported the psychometric properties of any of the versions of the SCAN;
2. Studies that used SCAN as a gold standard to validate other diagnostic and screening measures.

2.3. Data extraction

WF and AM screened the titles and abstracts of identified records. Two reviewers (WF and AM) then independently conducted full reviews of the articles. Discrepancies were resolved with discussion. The

excluded articles and reasons for exclusion were documented. The two reviewers divided the studies and extracted data, frequently discussing them when needed. The data extraction was started on May 6th 2024. The extraction sheet included: author, publication year, country, the SCAN sections assessed, participant characteristics and psychometric properties.

2.4. Risk of Bias Assessment

The two reviewers assessed bias using the COSMIN checklist [18]. The tool has 10 domains of psychometric properties: tool development/adaptation procedure, content validity, structural validity, internal consistency, cross-cultural validity, reliability, measurement error, criterion validity, hypotheses testing for construct validity, and responsiveness. Items under each domain are rated as very good, adequate, doubtful, and inadequate.

2.5. Data synthesis

We summarized the results of studies on SCAN's psychometric properties using descriptive statistics. We narrated the findings for domains where a statistical summary was not applicable. We narrated the profile of the studies that used SCAN as a gold standard, including the type of instrument, setting, and country.

3. Results

3.1. Study selection

See Fig. 1 for the PRISMA diagram. A total of 4,241 records were included. After title and abstract screening, 3,977 articles were removed because they did not study any of the SCAN versions. Two hundred ninety-six articles were considered for full-text review, including 32 articles from experts' suggestions and hand-searched for similar articles. Out of these, 203 were not eligible for this review because the studies mainly used SCAN for diagnosing neuropsychiatric problems for prevalence or intervention studies, rather than evaluating the properties of SCAN. In total, 93 articles were included in the review: 46 assessed the psychometric properties of SCAN, while the remaining 47 articles used SCAN as a gold standard.

3.2. Study characteristics

Among the 46 articles assessing the performance of SCAN, 31 articles were from high-income countries. Two were multi-county studies that included at least one low- and/or middle-income country [19, 20]. Out of the remaining 13, nine studies were from Thailand, two were from Ethiopia, one from Turkey and one from Brazil. In these studies, SCAN and PSE were translated into more than 19 languages. The publication year ranged from 1977 to 2011 (Fig. 2).

The study sample sizes ranged from eight video-recorded ratings to 1244 general population participants. These studies were conducted among participants admitted to hospitals, outpatient

departments, and general population settings. Some of the studies recruited participants from more than one place.

Twenty-eight studies reported on the full SCAN or PSE, while the remaining 18 studies reported on selected sections or items. Most of the studies used either trained psychiatrists, general practitioners or clinical psychologists, while some employed trainee psychiatrists, psychiatric nurses and lay interviewers (Table 1)

Table 1
Characteristics of studies examining the psychometric properties of SCAN or PSE

No	Author (Year)	Country (language)	SCAN sections	Rater
1.	Aboraya et al (1998) [53]	USA	SCAN 2·1	Psychiatrists
2.	Adamowski, Kiejna, Hadryś (2006) [27]	Poland (Polish)	SCAN 2·1	Psychiatrists
3.	Alem et al (2004) [39]	Ethiopia (Amharic)	SCAN 2·1	Psychiatrists
4.	Andrews et al (1995) [54]	Australia	Anxiety & depressive disorders	Clinical psychologists and psychiatrists
5.	Arunpongpaisal et al (2006) [55]	Thailand (Thai)	Alcohol and Use of Tobacco Section	Psychiatrists
6.	Brugha et al (1999) [41]	UK	SCAN 1	Psychiatrists & lay interviewers
7.	Brugha et al (1999) [56]	UK	Survey form	Psychiatrists & lay interviewers
8.	Brugha et al (2001) [49]	UK	SCAN	Psychiatrists
9.	Cheng et al (2001) [45]	China/Taiwan (Mandarin)	SCAN	Psychiatrists
10.	Compton et al (1996) [30]	USA	Substance use	Clinicians and senior psychiatric trainees
11.	Cooper et al (1977) [32]	UK	PSE-8	Psychiatrists vs trained non-psychiatrists (psychologist, sociologist, sociology student)
12.	Easton et al (1997) [34]	USA & Turkey (English & Turkish)	Alcohol & drug sections	Psychiatrists & psychologists
13.	Fink et al (2005) [28]	Denmark	SCAN 2·1	Neurologists
14.	Fink et al (2010) [29]	Denmark	SCAN 2·1	Physicians
15.	Hansen et al (2002) [57]	Denmark	SCAN 2·1	Psychiatrists
16.	Hanson 2001 [40]	Denmark	SCAN 2·1	Psychiatrists

No	Author (Year)	Country (language)	SCAN sections	Rater
17.	Hesselbrock et al (1999) [58]	UK	Substance use	Psychiatrists
18.	Krisanaprakornkit et al (2006) [59]	Thailand (Thai)	Mood disorders section	Psychiatrists
19.	Krisanaprakornkit et al (2007) [14]	Thailand (Thai)	Anxiety disorder section	Psychiatrists
20.	Lesage et al (1991) [60]	Canada (French)	PSE-9	Psychiatric nurses
21.	Luria & Berry (1979) [26]	USA	PSE	Psychiatrists & psychologists
22.	Małyszczak et al (2002) [31]	Poland	SCAN 2·1	Psychiatrists
23.	Maurer et al (1991) [5]	Switzerland	PSE-10	Not mentioned
24.	Mignolli et al (1988) [35]	Italy	PSE-9	Psychiatrists
25.	Nienhuis et al (2010) [33]	Netherlands	Mini-SCAN	Psychiatrists and clinical psychologists
26.	Paholpak et al (2006) [59]	Thailand	Somatoform and dissociative sections	Psychiatrists
27.	Paholpak et al (2008) [23]	Thailand	Psychotic disorder	Psychiatrists
28.	Paholpak et al (2010) [25]	Thailand	SCAN 2·1	Psychiatrists
29.	Patjanasoonorn et al (2011) [22]	Thailand	Eating disorder	Psychiatrists
30.	Pitta & Blay (1997) [61]	Brazil	Psychosis and anxiety	Psychiatrists
31.	Piyavhatkul et al (2008) [44]	Thailand	Cognitive	Psychiatrists
32.	Piyavhatkul et al (2008) [43]	Thailand	Stress and adjustment	Psychiatrists
33.	Pull et al (1997) [42]	Luxembourg, Greece and the US.	SCAN alcohol module	Psychiatrists
34.	Rijnders et al (2000) [6]	Netherlands	PSE-10	Psychologists

No	Author (Year)	Country (language)	SCAN sections	Rater
35.	Roca et al, (1999) [62]	Spain	PSE-10	Psychiatric trainees and psychologists
36.	Rodgers, Mann (1986) [38]	UK	PSE-9	Lay interviewer
37.	Schützwahl, Kallert, Jurjanz (2007) [13]	Germany	PSE-9:47 items from Part I of SCAN 2·1	Clinical psychologist, psychiatrist
38.	Schützwahl, Kallert, Jurjanz (2007) [21]	Dresden (Germany), Michalovce (Slovak Republic), Prague (Czech Republic), and Wrocław (Poland)	SCAN 2·1	Clinical psychologist, psychiatrist
39.	Shibre et al (2002) [37]	Ethiopia (Amharic)	SCAN 2·1	Psychiatrists
40.	Silverstone (1993) [63]	Canada	SCAN	Not mentioned
41.	Toft et al (2005) [64]	Denmark	SCAN 2·1	Physicians
42.	Toft et al (2010) [65]	Denmark	SCAN 2·1	General Practitioners
43.	Ustün et al (1997) [20]	Netherlands, Romania, Turkey, Greece, India, USA,, Nigeria, Luxembourg, Puerto Rico, Australia (2 cites)	Substance use	Not mentioned
44.	Van Hout Griez (1984) [66]	Netherland	PSE-9	Psychiatrists
45.	Vázquez-Barquero et al (1994) [19]	20 centers of different countries (Spanish)	PSE-10	Not mentioned
46.	Wing et al (1977) [3]	UK	PSE-9	Lay interviewer

3.3. Bias Assessment

Out of the ten COSMIN risk of bias checklist boxes, we did not rate the first two boxes: development and content validity, as none of the studies we included evaluated these domains for PSE or SCAN.

As most of the studies assessed only inter- and intra-rater reliability, we did not rate the studies for most of the boxes. We completed the checklist for studies that included information on Cronbach's alpha, test-retest reliability, concurrent validity, and structural validity. The most common issues we identified were low sample sizes, insufficient characterization of the sample, lack of reporting on the time interval between tests, and inadequate description of the comparator instrument used (Supplementary file 1).

3.4. Reliability of SCAN or PSE

Our review showed that different versions of the SCAN or PSE were assessed for their reliability, including internal consistency, inter-rater reliability, intra-rater reliability, and test-retest reliability.

In the studies that reported on the internal consistency of SCAN, Cronbach's alpha values ranged from 0.55 to 0.98 [13, 21]. This indicated good to excellent internal consistency, reflecting the coherence of different domains within SCAN and PSE.

There were also studies ($n = 8$) that examined the intra-rater reliability (instrument administered by the same rater at a prespecified interval, for example, two weeks) of SCAN. These studies reported acceptable agreement across various items or sections of SCAN.

The mean intra-rater kappa value ranged from substantial to almost perfect, although it was slightly lower (but still good) for some sections of SCAN. For somatoform and dissociative symptoms it was at 0.89, anxiety at 0.84, mood at 0.86, eating disorders at 0.76, use of alcohol at 0.82, stress and adjustment disorders at 0.94, psychosis at 0.76, cognitive impairment at 0.78, and observed behaviour, affect, and speech at 0.51. For stress and adjustment disorders and cognitive impairment, the values were between 0.65 and 0.86 [22–25]. The intra-class correlation coefficient (ICC) of the SCAN sections ranged from 0.84 to 0.86 [26].

The other reliability indicator was inter-rater reliability. Overall, there was acceptable agreement in most studies, although not across all settings and mental health conditions. One study indicated a significant level of difference in the parameters across the different mental health conditions: kappa for inter-rater reliability was 0.14 for neurotic disorders and 0.65 for psychotic disorders [27].

Section-level inter-rater agreement (kappa value) was also reported: somatoform or dissociative symptoms from 0.77 to 0.88 [28, 29], anxiety at 0.79, mood at 0.80, eating disorders at 0.73, use of alcohol at 0.66, stress and adjustment disorders at 0.90, psychosis at 0.68, cognitive impairment at 0.72, and observed behaviour, affect, and speech at 0.45 [22]. For the substance use disorders section of SCAN, the kappa level of agreement was 0.69 for alcohol, 0.49 for opioids, 0.61 for cocaine and 0.5 for cannabis [30].

A high level of inter-rater agreement was observed for the classification of schizophrenia and other psychotic disorders, while a lower level of agreement was found for other mental health conditions. For example, a Kappa value of 0.36 was found for depression, compared to 0.62 for schizophrenia [31]. One study reported an overall agreement of 0.81 from live interviews and 0.96 for tape-recorded interviews [3].

Six studies assessed the test-retest reliability of SCAN or PSE. One study indicated good consistency with a correlation coefficient of 0.67 [3]. The kappa values ranged from 0.7 to 0.9 [20], indicating substantial agreement. Another study reported a kappa level from 0.24 to 0.64 [6]. Test-retest also

worked better for psychotic symptoms ($\kappa = 0.82$) than non-psychotic symptoms ($\kappa = 0.16$) [32]. There was also a 0.90 κ for many sections of SCAN [33].

Regarding substance use disorders, there was acceptable consistency of the SCAN across time for alcohol (0.77–0.85), opiates (0.93–0.98), cannabis (0.63–0.78), and cocaine (0.69–0.74) [34]. One study indicated a Total Disagreement Index (TDI: total number of disagreements between raters/mean number of positive ratings), and the results were low, indicating the TDI mean is 0.09 (0–0.31) [35].

Validity of SCAN or PSE

Twenty-nine studies reported the concurrent, structural, technical, semantic or language validity of SCAN and PSE.

Concurrent validity of SCAN or PSE

Correlation between SCAN and clinical diagnoses using ICD-10 and DSM-IV has been found to vary, ranging from low to strong correlations [6, 20, 35–37]. It had strong or perfect agreement with a clinical diagnosis made by a nurse [38] or a trainee psychiatrist, especially for schizophrenia, but low agreement for depression [31, 39]. Another study showed that SCAN demonstrated 100% agreement for schizophrenia, 95.3% for bipolar disorder, and 93% for depression when compared to clinical diagnoses [39].

One study reported a sensitivity of 0.38 and specificity of 0.91 for physicians at admission and a sensitivity of 0.48 and specificity of 0.87 for nurses at admission. At discharge, the sensitivity was 0.40 and the specificity was 0.89 for physicians, while for nurses, the sensitivity was 0.56 and the specificity was 0.81 [40].

The sensitivity and specificity for the psychosis section of SCAN were notably higher at 0.69 and 0.94, respectively. However, specificity was lower for other modules, such as affective disorders (0.41) and neurotic disorders (0.5) [27]. The sensitivity against ICD-10 diagnoses from various studies ranged from 0.69 to 0.95, while specificity ranged from 0.41 to 0.94 [27, 41]

Construct validity of SCAN or PSE

Convergent validity between SCAN and mini-SCAN demonstrated substantial agreement [33, 42]. Structural validity, or the extent to which SCAN and PSE scores adequately reflect the dimensionality of the constructs they aim to measure, has also been examined using exploratory factor analysis.

For the PSE, the findings indicated a unidimensional structure. Many of the PSE items had high loading on the dominant factor [21], but the shared variance was small [38]. The first extracted factor accounted for 20.3% of the total variance, but the values of the second, third, and fourth factors accounted for only 5.5%, 4.5%, and 3.7%, respectively [38]. The exploratory factor analysis further indicated that SCAN effectively captures major psychiatric dimensions. In section 10, the mood disorder section factor

loading ranged from 0.88 to 0.97, the depression section with a factor loading ranged from 0.89 to 0.92, and the thinking section ranged from 0.24 to 0.88 [21]. There was also substantial agreement between the different sections of SCAN and its total scores [42].

Other properties

The SCAN and PSE have also demonstrated flexibility in administration, including duration of administration and practicality. The studies indicated that the duration of administration fluctuates based on the complexity of the condition and the specific section [35] being administered. The mini-SCAN interview takes about 25 minutes less than the full SCAN [33].

The alcohol module alone took about 40 minutes to administer, the stress-related disorders module took about 26 minutes [43], and 60 minutes for people with cognitive impairment [44]. On average, it took 57 minutes for somatoform and dissociative disorders [24].

In all the studies and all sections of SCAN, administering SCAN to people with impairment took longer compared to the healthy population. This was particularly the case among people with severe mental health conditions. Administering SCAN to people with schizophrenia took 140 ± 36.0 minutes, while it only took 82 ± 25.9 minutes (range 48–124 minutes for the healthy population [23]).

Two studies have indicated the acceptability of SCAN [33, 41]. Nienhuis et al. (2010) reported that 77% and 79% of the participants rated the SCAN and mini-SCAN interviews, respectively, as pleasant or very pleasant [33]. There were minimal reported difficulties in administering SCAN, especially when experience builds.

Studies that reported content, semantics, languages, and content validity indicated that SCAN could be effectively translated into different languages. There were no significant concerns reported (Table 2).

Table 2
Summary of psychometric properties of SCAN or PSE

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
Aboraya et al (1998)	kappa • 0.51 – 0.9 • Total item kappa for each interviewer (0.71 – 0.79)	-	-	-
Adamowski, Kiejna, Hadrys (2006)	Cohen's kappa (0.14 – 0.65) • Psychosis section (0.65) • Affective disorder (0.31) • Neurotic disorder (0.14)		Yule's Y • 0.57 – 0.71 Overall sensitivity (ICD 10 diagnosis) • 0.69 – 0.95 Specificity • 0.41 – 0.94 Psychosis section • Yule's Y (0.71) • Sensitivity (0.69) • Specificity (0.94) Affective disorder • kappa (0.31) • Sensitivity (0.95) • Specificity (0.41) Neurotic disorder • Sensitivity (0.95)	

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
			• Specificity (0·5)	
Alem et al (2004)			Kappa (Comparison of computer-assisted SCAN diagnoses and clinical diagnoses) • 100% for schizophrenia • 95·3% for bipolar • 93·0% for depression • 0·8% against the clinical diagnosis • Semantic validity established	
Andrews et al (1995)	Interrater • CIDI (1·00) • SCAN (0·67)			-
Arunpongpaisal et al (2006)	Inter-rater • Tobacco use disorder (0·84) • Alcohol use disorder (0·66) Intra-rater • Alcohol use disorder (0·82) • Tobacco use disorder (0·87)	-	-	-
Brugha et al (1999)	Inter-rater • Any ICD-10 non-psychotic diagnosis (0·25)	-	Overall sensitivity of CIS-R	-

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
	• All neurosis (0.26) • Any acute or chronic case of depression (0.13)		• 0.5 Specificity • 0.8 – 0.9	
Brugha et al (1999)	-	-	Concordance for any disorder • 0.74 (0.57, 0.91) Any specific psychotic disorder • 0.63 (0.40, 0.86) For any specific neurotic disorder • 0.63 (0.43, 0.83). Sensitivity (ICD-10) • 0.6-0.82 Specificity • 0.83 – 0.93 • No evidence of rater bias	• At first, lay interviewers found SCAN quite daunting • With experience in carrying out interviews, their confidence grew and with it enthusiasm for the procedure • The SCAN trainers were happy with the progress of the lay interviewers
Brugha et al (2001)			Concordance of SCAN and CIDI • 0.03 – 0.48 when the CIDI preceded the SCAN • 0.33 when it followed SCAN	• Mean duration of the SCAN, 34 min (15 to 75).
Cheng et al. (2001)	Agreement between UK & US	-	-	-

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
	vs Taiwanese psychiatrists • 69% for section 8 • 100% for section 11			
Compton et al. (1996)	Diagnosis agreement • Alcohol (0·69) • Opiate (0·49) • Cocaine (0·61) • Cannabis (0·5)	-	-	-
Cooper et al. (1977)	Overall interrater agreement was 0.77 • Situational anxiety (0·34) • Depersonalization (0·96) • Item level agreement of the 150 items (0·74)	Test re-test (0.49) • Worry (0·16) • Relationships and ideas of reference (0·82)	-	-
Easton et al. (1997)	-	ICD 10 dependence • Alcohol (0·77 – 0·85) • Opiates (0·93 – 0·98) • Cannabis (0·63 – 0·78) • Cocaine (0·69 – 0·74)	-	-
Fink et al (2005)	• Interrater agreement of kappa = 0·88	-	-	-

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
Fink et al (2010)	• Inter-rater agreement on the presence of/lack of a somatoform diagnosis (kappa = 0.82) among six interviewers	-	•	-
Hansen et al (2001)	• Interrater agreement kappa (psychiatrists) = 0.88	-	Diagnosis by physicians and nurses at Internal medicine inpatients vs SCAN diagnosis • Physicians: sensitivity (0.38) and specificity 0.91 at admission • Nurses: sensitivity (0.48) and specificity (0.87) at admission • Physicians: sensitivity (0.40) and specificity (0.89) at discharge • Nurses: sensitivity (0.56) and specificity (0.81) at discharge	-
Hansen et al (2002)	• Interrater agreement was kappa = 0.88 (agreement on 16 of 17)	-	-	-
Hesselbrock et al (1999)	SSAGA comparison with the SCAN • Alcohol dependence (0.63)	-	-	-

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
	• Sedative dependence (0.48) • Cannabis dependence (0.53) • Stimulant dependence (0.85) • Opioid dependence (0.73) • Major depression (0.71)			
Krisanaprakornkit et al (2006)	Intra-rater • Sections 6 (0.80) • section 7 (0.67) • Section 8 (0.78) • Section 10 (0.84)	-	• Some adaptations were made to words or sequences of sentences describing symptoms to make them more understandable in the Thai cultural and linguistic context	-
Krisanaprakornkit et al (2007)	Overall inter-rater reliability • (0.79) Section 3 • (0.75) Section 4 • (0.8) Section 5 • (0.76) Intra-rater reliability (two weeks) • (0.84)	-	• Some adaptations were made to words or sequences of sentences describing symptoms to make them more understandable in the Thai cultural and linguistic context	-

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
Lesage et al (1991)	Inter-rater • 0.75 Positive agreement rate • 0.59 – 0.69	-	-	-
Luria & Berry (1979)	Inter-rater • ICC (0.84 – 0.86)	-	-	-
Małyszczak et al (2002)	Cohen's kappa • Schizophrenia (0.62) Depression (0.36)	-	Kappa coefficient • Block F2 (0.78) • Schizophrenia (0.62) • Block F3 (0.42) • Depression (0.36) • Low agreement between clinical and SCAN diagnosis, particularly for mood disorder	-
Mignolli et al (1988)	The total disagreement index (TDI) • 0.09 (0–0.31) Mean serious disagreement index (SDI) • 0.04 (0–0.21) • Between interviewer and reinterviewed were 0.37 (0–1.05) and 0.21 (range 0–0.86) respectively	-	-	-

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
Nienhuis et al (2010)	-	Prevalence SCAN vs Mini-SCAN • No disorder (15 vs 15) • Affective disorder (30 vs 32) • Affective psychosis (12 vs 11) • Anxiety disorder (34 vs 33) • non-affective psychosis (15 vs 15) Mini-SCAN • Sensitivity (0.67 – 0.87) • Specificity (0.92 – 0.97) • PPV (0.73 – 0.97) • NPV (0.96 – 0.97)	Concurrent validity of the mini-SCAN • 0.80	• Mini-SCAN interviews were 25 minutes shorter than the SCAN • SCAN, 77% rated the interview as pleasant or very pleasant • For the mini-SCAN this percentage was 79%.
Paholpak et al (2006)	Kappa (inter-rater) • 0.81 – 1.0, 0.61 – 0.80 and 0.00 – 0.20 in 49.6%, 30.0% and 8.9% of the items respectively • Kappas could not be calculated for 11.5% of the items The intra-rater reliabilities were	-	• Semantic validity established	• Averaged = 57.1 ± 12.1 minutes while it was 42.1 ± 13.9 minutes for a normal subject

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
	• 0.81 – 1.0, 0.61 – 0.80 and 0.00– 0.20 in 54.9%, 26.5% and 2.7% of the items respectively • Kappas could not be calculated for 15.9% of the Items.			
Paholpak et al (2006)	Kappa (inter and intra-rater) agreement • Somatoform and dissociative symptoms module (0.77, 0.85) • Anxiety (0.79, 0.84) • Mood (0.80, 0.86) • Eating disorders (0.73, 0.76) • Use of alcohol (0.66, 0.82) • Stress and adjustment disorders (0.90, 0.94) • Psychosis (0.68, 0.76) • Cognitive impairment (0.72, 0.78) • Observed behavior, affect and speech module (0.45, 0.51)	-	• Semantic validity established	-

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
Patjanasoonporn et al (2011)	• Kappa of the eating disorder section 0.73, (0.68,0.77). • Mean intra-rater reliabilities (0.76; 0.70-0.82) • 68.4% & 31.58% of the items had a substantial and almost perfect kappa respectively	-	• Semantic validity established	-
Pitta & Blay (1997)	-	-	Kappa between ICD-9 'hysteria' and 'other reactive and non-specified psychoses' and the corresponding categories and the PSE/CATEGO program - 0.03 and 0.19	-
Piyavhatkul et al (2008)	• Kappa (0.90 (SD = 0.12) • Item level agreement 77.05%- 85.26%	-	• Semantic validity established	• Averaged 17.92 minutes (25.59 for people with stress-related disorders and 6.41 for normal subjects).
Piyavhatkul et al (2008)	• Intra-rater: mean 0.94 (SD = 0.09) • Perfect level among many items; however, three items (3.66%) had poor and nine (6.67%) only slight inter-rater agreement • Inter and intra-rater reliability of the continuous	-	• Semantic validity established	• Cognitive impairment or Decline Section averaged 48.99 minutes (59.71 for participants with cognitive impairment and 33.77 for normal participants)

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
	data was 0.93 and 0.96.			
Piyavhatkul et al (2008)	Mean & SD of inter-rater reliability for sections 6, 17, 18 and 19 • 0.66 ± 0.17 • 0.71 ± 0.16 • 0.70 ± 0.22 • 0.64 ± 0.23 Intra-rater reliability • 0.65 ± 0.11 • 0.74 ± 0.17 • 0.86 ± 0.17 • 0.80 ± 0.18 • Some sections had items with 100 percent agreement from the same rater when rated 2 weeks apart.	-	• Semantic validity established	People with schizophrenia • 140.2 ± 36.0 minutes (75–193) Comparison • 81.9 ± 25.9 minutes (48–124)
Pull et al (1997)		-	• The SCAN and AUDADIS-ADR also agreed more closely to one another than to the CIDI in terms of the neglect of interests and the much time spent and continued use despite physical and/or psychological harm criteria.	• 40.3 min to administer

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
Rijnders et al (2000)	Kappa • 0.56	• 0.24 – 0.64	• Sensitivity as well as specificity proved to be substantial to almost perfect. • The agreement per interviewer about the reference diagnoses ranged from 87% (diagnostic group) to 94% (diagnostic caseness). • Agreement on the syndrome level (without duration and interference criteria of DSM-IV) was excellent.	-
Roca et al, (1999)	-	-	• Of subjects with a positive GHQ, 47% were considered non-cases" on the SCAN using the ICD-10 criteria	-
Rodgers, Mann (1986)	-	-	The first factor extracted accounted • 20.3% of the total variance Second, third, and fourth factors • 5.5, 4.5 and 3.7% respectively • 71.5% showed perfect agreement and	-

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
			24·9% had only 1 or 2 inconsistencies. • The remaining 3·6% contained 3 or more instances of disagreement between the nurse and the standard rating	
Schützwahl, Kallert, Jurjanz (2007)	Cronbach α • 0·55 to 0·98*	-	One factor solution Section 10-mood • highest loading-0·88 – 0·97 Depression section • 0·890 to 0·92 and Section 7 thinking • 0·24 to 0·88 Acceptable face validity	-
Schützwahl, Kallert, Jurjanz (2007)	Cronbach α • 0·50 – 0·70 for most scale and > 0·7 for a few scales	-	High correlation between the scales generated by PCA and the section-specific scales. SCAN sub-sections	-
Shibre et al (2002)	• Inter-rater reliability of the SCAN was not assessed formally, but a total agreement among the interviewers in terms of SCAN	-	Overlap between SCAN and CIDI • 26% Semantic validity established	-

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
	cases/non-case as well as in terms of diagnostic category.			
Silverstone (1993)	-	-	• Significant overall correlations for DSM-III–R diagnoses between SCAN and SADS • There were no significant differences in the incidence of diagnosis for any of the diagnostic groups	-
Toft et al (2005)	• Interrater agreement (kappa = 0.86)	-	-	-
Toft et al (2010)	• Interrater agreement (kappa 0.82)	-	-	-
Ustün et al (1997)	-	Kappa • 0.7 – 0.9	Correlation with ICD-10 and DSM-IV • Ranges from 0.36 – 0.68.	-
Van Hout Griez (1984)	-	-	Correlation between subsections of PSE and Wolpe and Lazarus' Serf Rating Scale for Assertiveness, fear Survey Schedule, Zung's Self Rating Scale for Depression and Maudsley Obsessive Compulsive Inventory	• Average 30 mins

Author and Year	Internal consistency*, intra-rater and Inter-rater reliability	Test-retest reliability	Validity	Feasibility/duration of administration and acceptability
			• 0.24 – 0.67.	
Wing et al (1977)	• For tape-recorded interviews, r (0.96) • Overall agreement for all cases is (0.81)	• r (0.67)	• Two sets of total PSE scores (0.73)	-

3.5. SCAN or PSE as a gold standard

Forty-seven articles used SCAN or PSE versions to evaluate the psychometric properties of other diagnostic or screening tools. The publication year ranged from 1989 to 2023. Most of the studies, except for five (Ethiopia (n = 3), Nigeria, and South Africa), were conducted in high-income countries.

A total of 31 tools were validated using SCAN or PSE as a gold standard. Of the 31, 18 were used to assess depression, while the other measures were used to assess symptoms of psychosis, anxiety, somatoform, and substance use problems. Some of the scales or assessment instruments (n = 4), such as the CIDI, were used to identify more than one psychiatric problem. These 31 measures were validated in more than 15 different languages (Supplementary file 2).

4. Discussion

We aimed to systematically review the psychometric properties of the various revisions of the SCAN. We found 46 original psychometric papers on SCAN and PSE and another 47 articles that used SCAN as a gold standard to evaluate the psychometric properties of other mental health measures. Most of the studies used multiple sections of SCAN, while some used one or two sections out of the 27 sections.

The internal consistency of the items in SCAN and PSE ranged from good to excellent. Additionally, both modules demonstrated acceptable intra-rater, inter-rater, and test-retest reliability. However, some studies reported percentages and mean [30, 44, 45] to calculate various types of reliabilities. This might introduce methodological biases and require cautious interpretation.

There was also evidence supporting the concurrent, semantic, and content validity of both SCAN and PSE, though there was no evidence for predictive validity. The good to excellent internal consistency of SCAN and PSE suggests that the items effectively measure the underlying construct of psychopathology. But, internal consistency should have been calculated for each section of the SCAN or PSE, not the total score [46]. Another issue to consider is that the high internal consistency of the SCAN

and PSE might be attributed to the large number of items in these scales, rather than their actual reliability [47]. High internal consistency is not an ideal indicator for tools with many items [46].

We also found acceptable agreement across various items or sections of SCAN and PSE. However, methodological concerns arise from the use of percentage agreement and Cohen's Kappa when the studies calculated test-retest, intra-rater, and inter-rater reliability. Unlike Kappa, percentage agreement does not account for the possibility that agreement may occur by chance.

Additionally, high agreement among raters on specific sections, particularly those related to schizophrenia or psychotic disorders, could positively influence the overall percentage of agreement when considering the entire SCAN and PSE. Only a few studies calculated the ICC, which is a rigorous methodology that considers total variance, including between participants, between raters, and residual error. However, these studies did not report measurement errors [46].

Although the evidence is not robust, both SCAN and PSE appear more reliable for psychotic disorders, which present with more severe, clear, and observable symptoms compared to neurotic or common mental disorders. Of interest, the strengths and weaknesses of the SCAN and instruments such as CIDI are inverse, as CIDI does not perform well for psychotic disorders [48]. This suggests that improving the sections on common mental disorders in the SCAN could be a focus for future research.

As the PSE and SCAN have been translated into many languages, considerable experience has been gained. Translating concepts, not just words, and the flexibility of SCAN, allowing administrators to explain and probe, suggests its strong semantic validity [4]. Studies also showed that they were able to establish that the back-translation of SCAN and PSE into the original English is appropriate.

Because SCAN is typically considered a gold standard, criterion validity has been difficult to study. Few studies have reported acceptable sensitivity and specificity of SCAN in comparison to diagnosis by clinicians, particularly for psychotic disorders as opposed to affective disorders [27].

The positive and negative predictive power of SCAN remains uncertain, as it has not been extensively studied. Correlations between SCAN and diagnosis using ICD-10 and DSM-IV were found to vary, ranging from low to strong correlation [6, 20, 36, 37].

The semi-structured nature of SCAN allows flexibility for the test administrator to rephrase and probe, and the common practice of its use by highly trained mental health professionals could potentially contribute to better validity compared to other structured diagnostic tools, such as the CIDI [49].

There is a need for strong studies that report the construct validity of SCAN using factor analysis. There is only such a study which reported one dominant factor and high factor loading of the items [21]. This finding is like studies on CIDI and other diagnostic tools, which tend to be unidimensional per section [50]. Further research on the cultural invariance of SCAN is warranted, as this would facilitate cross-cultural comparisons of psychiatric disorders.

SCAN is accepted and used globally [8, 9, 51, 52]. Previous studies have also noted that SCAN can be effectively utilized by trained and experienced lay interviewers [41]. The computer algorithm for interpretation and its dimensional measurement capabilities are additional strengths of SCAN [2].

Conclusion

Considering the methodological weakness of the studies that reported the psychometric properties of SCAN when evaluated against recently recommended standards, there is a clear need for further studies. Additionally, we couldn't find recent studies that evaluate the psychometric properties of SCAN.

Despite these limitations, SCAN continues to be a promising tool in terms of some of its psychometric properties and practical feasibility for diagnosing psychiatric disorders. The established semantics, content, along with some accepted reliability indicators, underscore its potential as a valuable diagnostic instrument.

The next version of SCAN must consider its potential limitations, particularly in the non-psychotic sections. Additionally, it is important to revise the current version considering the ICD-11 and DSM-5 diagnostic systems, as the current version is based on the previous versions.

Contributions

MO, AF, AA, and WF conceived the study. WF, AM, AF, and AA contributed to the design of the review and the protocol development. WF and AM search the literature, data collection, and quality appraisal. All authors participated in contacting researchers who used SCAN and in retrieving some of the old publications. WF and AM drafted the manuscript. All authors revised the manuscript and approved the final manuscript.

Declarations

Contributions

MO, AF, AA, and WF conceived the study. WF, AM, AF, and AA contributed to the design of the review and the protocol development. WF and AM search the literature, data collection, and quality appraisal. All authors participated in contacting researchers who used SCAN and in retrieving some of the old publications. WF and AM drafted the manuscript. All authors revised the manuscript and approved the final manuscript.

Availability of Data and Materials

The authors confirm that the data supporting the findings of this study are available in the article and its supplementary materials.

Declarations of interest

MVO is currently an employee of WHO, while SC is a former employee of the organization.

Financial declarations

WF and AM received support from WHO. WF and CH receive support from the Wellcome Trust through grants 222154/Z20/Z. CH receives support from NIHR through the NIHR Global Health Research Group on Homelessness and Mental Health in Africa (NIHR134325), using UK aid from the UK Government. CH also receives support from WT grant 223615/Z/21/Z. AM receives support from DELTAS Africa Initiative through the African Mental Health Research Initiative (AMARI) II project. The DELTAS Africa Initiative is a programme of the Science for Africa Foundation being implemented with support from Wellcome and the UK Foreign Commonwealth and Development Office (FCDO). The views expressed are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care or WHO.

For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Author Manuscript version arising from this submission.

References

1. Sheehan DV et al (1998) The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 59(20):22–33
2. WHO (1994) Schedules for clinical assessment in neuropsychiatry: version 2. American Psychiatric
3. Wing J et al (1977) Reliability of the PSE used in a population study. *Psychol Med* 7(3):505–516
4. Wing JK et al (1990) SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Arch Gen Psychiatry* 47(6):589–593
5. Maurer K et al (1985) *[Experiences with Schedules for Clinical Assessment in Neuropsychiatry (SCAN) within the scope of a multicenter field study]*. *Schweiz Arch Neurol Psychiatr* 142(3): pp. 235 – 45
6. Rijnders CA et al (2000) Psychometric properties of the schedules for clinical assessment in neuropsychiatry (SCAN-2.1). *Soc Psychiatry Psychiatr Epidemiol* 35(8):348–352
7. Morgan C et al (2023) Epidemiology of untreated psychoses in 3 diverse settings in the global south: The international research program on psychotic disorders in diverse settings (INTREPID II). *JAMA psychiatry* 80(1):40–48
8. Hanlon C et al (2024) Studying the context of psychoses to improve outcomes in Ethiopia (SCOPE): Protocol paper. *PLoS ONE* 19(5):e0293232

9. Hanlon C et al (2016) Task sharing for the care of severe mental disorders in a low-income country (TaSCS): study protocol for a randomised, controlled, non-inferiority trial. *Trials* 17:1–14
10. Uscamayta Ayvar M et al (2022) *Non-affective psychosis in traditional Andean culture*. *Transcult Psychiatry*, : p. 13634615221099795
11. Qassem T et al (2015) Prevalence of psychosis in black ethnic minorities in Britain: analysis based on three national surveys. *Soc Psychiatry Psychiatr Epidemiol* 50(7):1057–1064
12. de Erausquin GA et al (2022) *Chronic neuropsychiatric sequelae of SARS-CoV-2: Protocol and methods from the Alzheimer's Association Global Consortium*. *Alzheimers Dement (N Y)*, 8(1): p. e12348
13. Schützwohl M, Kallert T, Jurjanz L (2007) Psychometric properties of dimensional measures derived from the latest German version of the Schedules for Clinical Assessment in Neuropsychiatry 2.1]. *Nervenarzt* 78(3):304–308
14. Krisanaprakornkit T et al (2007) The validity and reliability of the WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN Thai Version): anxiety disorders section. *J Med Assoc Thai* 90(2):341–347
15. Page MJ et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:n71
16. Prinsen CAC et al (2018) COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 27(5):1147–1157
17. Gagnier JJ et al (2021) COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Qual Life Res* 30(8):2197–2218
18. Mokkink LB et al (2018) COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res* 27(5):1171–1179
19. Vázquez-Barquero JL et al (1994) [Development and verification of the Spanish version of the scanning system psychiatric interview (Questionnaires for clinical evaluation in neuropsychiatry)]. *Actas Luso Esp Neurol Psiquiatr Cienc Afines* 22(3):109–120
20. Ustün B et al (1997) WHO Study on the reliability and validity of the alcohol and drug use disorder instruments: overview of methods and results. *Drug Alcohol Depend* 47(3):161–169
21. Schützwohl M, Kallert T, Jurjanz L (2007) Using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN 2.1) as a diagnostic interview providing dimensional measures: cross-national findings on the psychometric properties of psychopathology scales. *Eur Psychiatry* 22(4):229–238
22. Patjanasoonporn N, Paholpak S, Krisanaprakornkit T (2011) Validity and reliability study of the Thai version of WHO Schedules for Clinical Assessment in Neuropsychiatry: Eating Disorders Section. *J Med Assoc Thai* 94(4):490–497
23. Paholpak S et al (2008) Validity and reliability study of the Thai version of WHO schedules for clinical assessment in neuropsychiatry: sections on psychotic disorders. *J Med Assoc Thai* 91(3):408–416

24. Paholpak S et al (2006) Validity and reliability study of the Thai version of WHO SCAN: somatoform and dissociative symptoms section. *J Med Assoc Thai* 89(4):473–483
25. Paholpak S et al (2010) Validity and reliability study of the Thai version of WHO Schedules for Clinical Assessment in Neuropsychiatry version 2.1. *J Med Assoc Thai* 93(4):497–501
26. Luria RE, Berry R (1979) Reliability and descriptive validity of PSE syndromes. *Arch Gen Psychiatry* 36(11):1187–1195
27. Adamowski T, Kiejna A, Hadryś T (2006) Study of compatibility of psychiatric diagnoses with ICD-10 diagnostic criteria using the SCAN questionnaire]. *Psychiatr Pol* 40(4):761–773
28. Fink P, Hansen MS, Søndergaard L (2005) Somatoform disorders among first-time referrals to a neurology service. *Psychosomatics* 46(6):540–548
29. Fink P, Ørnbøl E, Christensen KS (2010) The outcome of health anxiety in primary care. A two-year follow-up study on health care costs and self-rated health. *PLoS ONE* 5(3):e9873
30. Compton WM et al (1996) Comparing assessments of DSM-IV substance dependence disorders using CIDI-SAM and SCAN. *Drug Alcohol Depend* 41(3):179–187
31. Małyszczak K et al (2002) [Comparison between a SCAN diagnosis and a clinical diagnosis]. *Psychiatr Pol* 36(6 Suppl):377–380
32. Cooper JE et al (1977) Further studies on interviewer training and inter-rater reliability of the Present State Examination (PSE). *Psychol Med* 7(3):517–523
33. Nienhuis FJ et al (2010) Validity of a short clinical interview for psychiatric diagnosis: the mini-SCAN. *Br J Psychiatry* 196(1):64–68
34. Easton C et al (1997) Test-retest reliability of the alcohol and drug use disorder sections of the schedules for clinical assessment in neuropsychiatry (SCAN). *Drug Alcohol Depend* 47(3):187–194
35. Mignolli G et al (1988) Inter-rater reliability of PSE-9 (full version): an Italian study. *Soc Psychiatry Psychiatr Epidemiol* 23:30–35
36. Fink P et al (1995) Psychiatric morbidity in primary public health care: a Nordic multicentre investigation. Part I: method and prevalence of psychiatric morbidity. *Acta Psychiatr Scand* 92(6):409–418
37. Shibre T et al (2002) An evaluation of two screening methods to identify cases with schizophrenia and affective disorders in a community survey in rural Ethiopia. *Int J Soc Psychiatry* 48(3):200–208
38. Rodgers B, Mann SA (1986) The reliability and validity of PSE assessments by lay interviewers: a national population survey. *Psychol Med* 16(3):689–700
39. Alem A et al (2004) Comparison of computer assisted scan diagnoses and clinical diagnoses of major mental disorders in Butajira, rural Ethiopia. *Ethiop Med J* 42(2):137–143
40. Hansen MS et al (2001) Mental disorders among internal medical inpatients: prevalence, detection, and treatment status. *J Psychosom Res* 50(4):199–204
41. Brugha TS et al (1999) Cross validation of a general population survey diagnostic interview: a comparison of CIS-R with SCAN ICD-10 diagnostic categories. *Psychol Med* 29(5):1029–1042

42. Pull CB et al (1997) Concordance between ICD-10 alcohol and drug use disorder criteria and diagnoses as measured by the AUDADIS-ADR, CIDI and SCAN: results of a cross-national study. *Drug Alcohol Depend* 47(3):207–216
43. Piyavhatkul N, Krisanaprakornkit T, Paholpak S (2008) Stress and Adjustment Disorder Section of WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN)–Thai version: validity and reliability study. *J Med Assoc Thai* 91(8):1263–1268
44. Piyavhatkul N et al (2008) Validity and reliability of WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN)-Thai version: Cognitive Impairment or Decline Section. *J Med Assoc Thai* 91(7):1129–1136
45. Cheng AT et al (2001) Cross-cultural implementation of a Chinese version of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) in Taiwan. *Br J Psychiatry* 178:567–572
46. Mokkink LB et al (2012) COSMIN checklist manual. University Medical Center, Amsterdam
47. Meyer GJ et al (2001) Psychological testing and psychological assessment. A review of evidence and issues. *Am Psychol* 56(2):128–165
48. Cooper L, Peters L, Andrews G (1998) Validity of the Composite International Diagnostic Interview (CIDI) psychosis module in a psychiatric setting. *J Psychiatr Res* 32(6):361–368
49. Brugha TS et al (2001) A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychol Med* 31(6):1001–1013
50. Gelaye B et al (2013) Diagnostic validity of the composite international diagnostic interview (CIDI) depression module in an East African population. *Int J Psychiatry Med* 46(4):387–405
51. Gureje O et al (2020) Effect of collaborative care between traditional and faith healers and primary health-care workers on psychosis outcomes in Nigeria and Ghana (COSIMPO): a cluster randomised controlled trial. *Lancet* 396(10251):612–622
52. Roberts T et al (2020) INTREPID II: protocol for a multistudy programme of research on untreated psychosis in India, Nigeria and Trinidad. *BMJ open* 10(6):e039004
53. Aboraya A et al (1998) Schedules for Clinical Assessment in Neuropsychiatry (SCAN): introduction to WV's mental health community. *W V Med J* 94(6):326–328
54. Andrews G et al (1995) A comparison of two structured diagnostic interviews: CIDI and SCAN. *N Z J Psychiatry* 29(1):124–132 Aust
55. Arunpongpaisal S et al (2006) Inter- and intra-rater reliability of the Thai version of SCAN: Use of Alcohol and Use of Tobacco Section. *J Med Assoc Thai* 89(12):2129–2137
56. Brugha TS et al (1999) The survey form of SCAN: the feasibility of using experienced lay survey interviewers to administer a semi-structured systematic clinical assessment of psychotic and non-psychotic disorders. *Psychol Med* 29(3):703–711
57. Hansen MS et al (2002) Use of health services, mental illness, and self-rated disability and health in medical inpatients. *Psychosom Med* 64(4):668–675

58. Hesselbrock M et al (1999) A validity study of the SSAGA-a comparison with the SCAN. *Addiction* 94(9):1361–1370
59. Krisanaprakornkit T, Paholpak S, Piyavhatkul N (2006) The validity and reliability of the WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN Thai Version): Mood Disorders Section. *J Med Assoc Thai* 89(2):205–211
60. Lesage AD, Cyr M, Toupin J (1991) Reliable use of the Present State Examination by psychiatric nurses for clinical studies of psychotic and nonpsychotic patients. *Acta psychiatrica Scandinavica* 83(2):121–124
61. Pitta JC, Blay SL (1997) *Psychogenic (reactive) and hysterical psychoses: a cross-system reliability study*, in *Acta psychiatrica Scandinavica*. United States, pp 112–118
62. Roca M et al (1999) Mental disorders on the island of Formentera: prevalence in general population using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Soc Psychiatry Psychiatr Epidemiol* 34(8):410–415
63. Silverstone PH (1993) SCAN accurately gives diagnoses according to DSM-III–R criteria: Validation in psychiatric patients using SADS. *International Journal of Methods in Psychiatric Research*
64. Toft T et al (2005) Mental disorders in primary care: prevalence and co-morbidity among disorders. results from the functional illness in primary care (FIP) study. *Psychol Med* 35(8):1175–1184
65. Toft T et al (2010) Training general practitioners in the treatment of functional somatic symptoms: effects on patient health in a cluster-randomised controlled trial (the Functional Illness in Primary Care study). *Psychother Psychosom* 79(4):227–237
66. van den Hout MA, Griez E (1984) Validity and utility of the present state examination in assessing neurosis: empirical findings and critical considerations. *J Psychiatr Res* 18(2):161–172

Figures

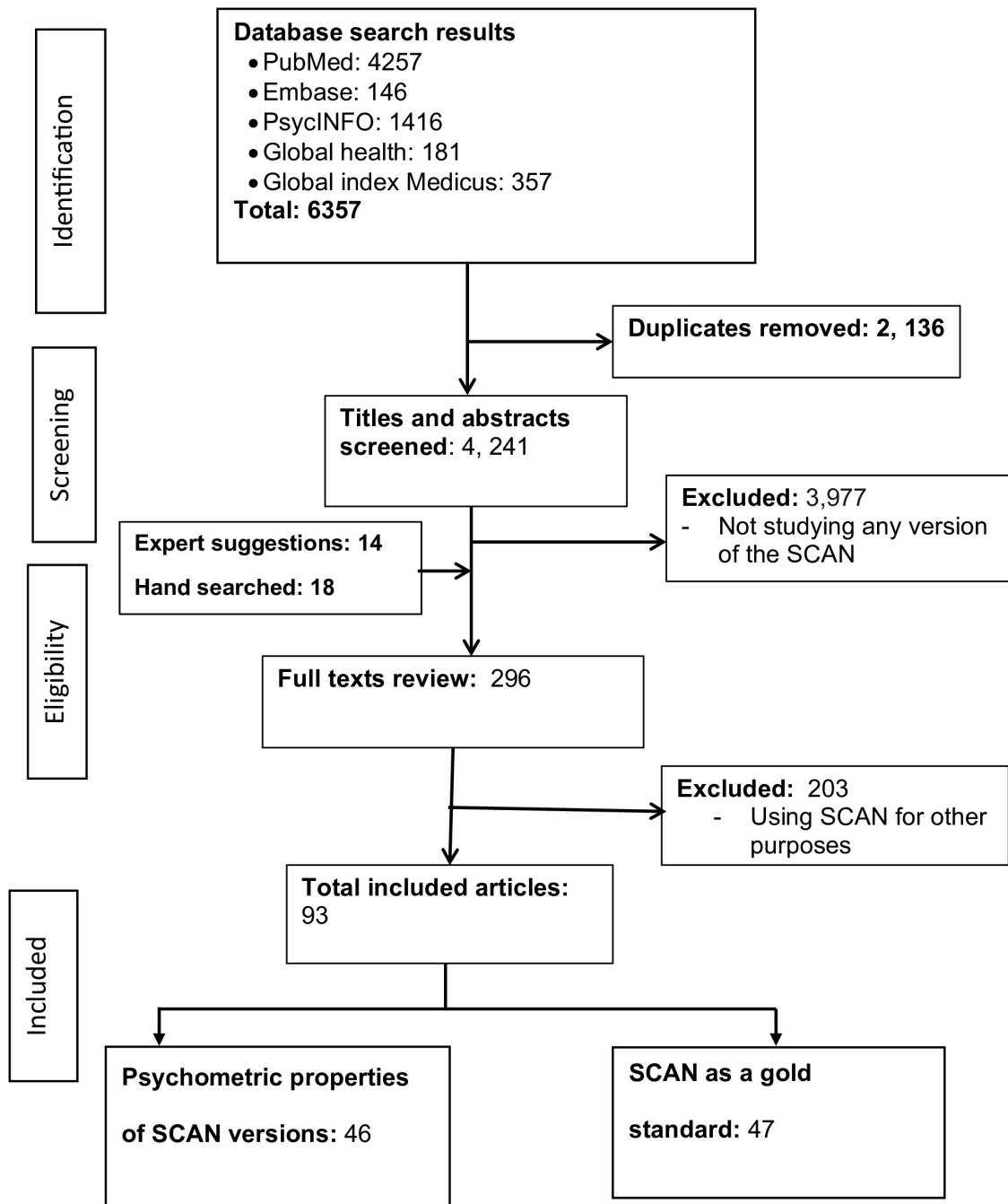
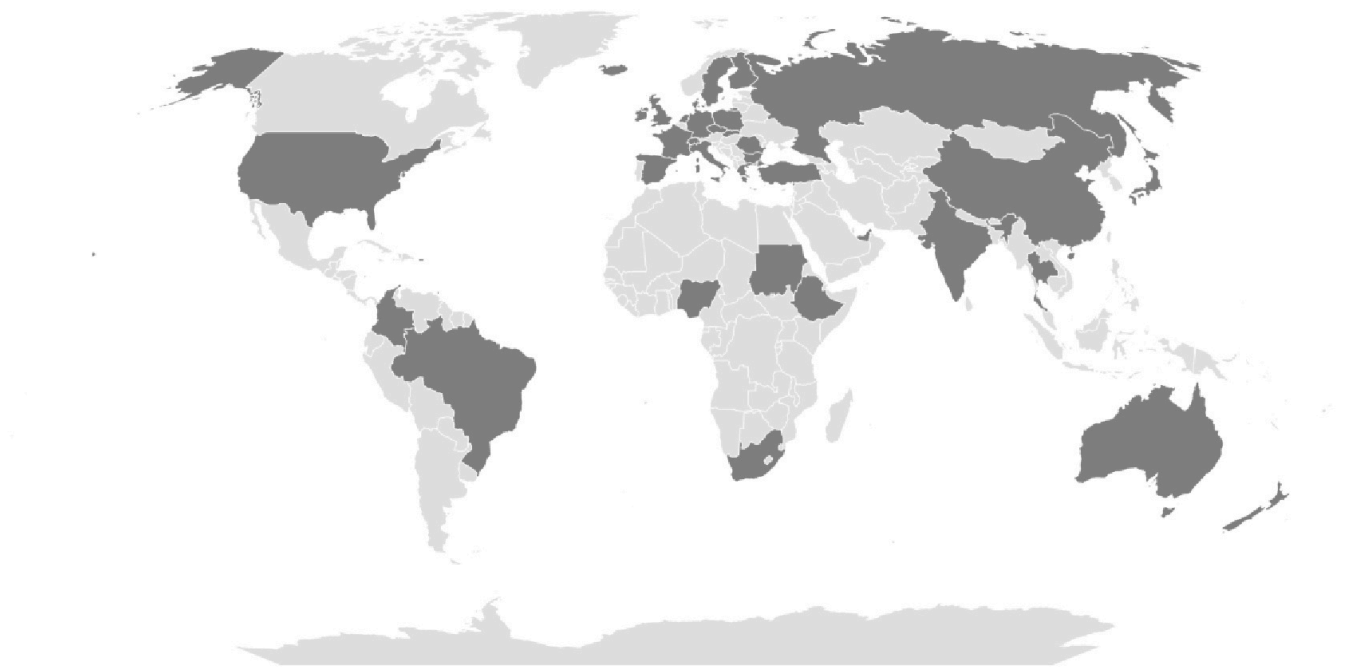


Figure 1: PRISMA flow diagram of the study selection process

Figure 1

See image above for figure legend



© Australian Bureau of Statistics, GeoNames, Microsoft, Navinfo, Open Places, OpenStreetMap, Overture Maps Fundation, TomTom, Zenrin
Powered by Bing

Figure 1: Countries where SCAN has been used

Figure 2

See image above for figure legend

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryfileSearchstrategy.docx](#)