Early detection

Introduction

The need for comprehensive evaluation of clinical assessment instruments and criteria for the early detection of schizophrenia is motivated by clinical and policy making considerations. The importance of early detection and intervention of schizophrenia has arisen from observations that the longer the period of untreated psychosis, the worse the outcomes. Therefore investment of resources in the early stages prior to the onset of psychotic symptoms might substantially reduce the frequency and severity of chronic disablement.

Early detection in this context may be defined as the identification of people thought to be at “ultra-high risk” of developing psychosis - those displaying “at-risk mental states”, traditionally referred to as “prodromal symptoms.” A number of new assessment instruments have been constructed aimed at detecting at-risk mental states.

Two main approaches have been adopted\(^1\). One focuses on a triad of at risk mental states in individuals defined as: (1) a Family History (FH) of psychosis plus non-specific symptoms and recent decline in functioning, (2) recent onset Attenuated Psychotic Symptoms (APS) with decline in functioning and (3) Brief Limited Intermittent Psychotic Symptoms (BLIPS) with decline in functioning. The other is based on Huber’s Basic Symptoms (BS) that focuses on a detailed way of describing phenomenological (subjective) disturbances.

Three scales of measurement in the former approach are: the Comprehensive Assessment of At-Risk Mental States (CAARMS), the Structured Interview of Prodromal Syndromes (SIPS), and the Basel Screening Instrument for Psychosis (BSIP). There are also differences in the cut-off criteria; for example, some subjects that were classified in the Brief Limited Intermittent Psychotic Symptoms (BLIPS) subgroup by CAARMS would be classified as psychotic by SIPS, and some subjects classified in the APS subgroup by CAARMS would be regarded as BLIPS by SIPS. SIPS dictates a lower threshold for a psychosis classification. The symptoms highlighted by various instruments also differ, for example, CAARMS emphasizes more stable and persistent symptoms pertaining to the FH subgroup, whereas SIPS emphasizes more recent symptoms, reflecting any deterioration.

Two Basic Symptoms scales were examined, the Bonn Scale for the Assessment of Basic Symptoms (BSABS), and the Schizophrenia Prediction Instrument (SPI-A [adult] and SPI-CY [child and youth]). The BSABS was designed to detect early subtle symptoms in the domains of perception, cognition, language, motor function, will, initiative and level of energy, and stress tolerance. The SPI-A was designed to be used as a supplement to SIPS and CAARMS. Ratings are linked to cognitive deficits that may be present prior to a prodromal state. Symptoms rated as severe on SPI-A should correspond to symptoms of moderate severity on SIPS, symptoms rated as severe on SIPS should correspond to moderate ratings on PANSS. It is important to determine which scales are most sensitive to detecting these symptoms.

Screening instruments have also been developed for research recruitment purposes, and for clinical screening\(^2\). PRODscreen is intended to detect persons with elevated risk of psychosis, who will subsequently be assessed with SIPS. PRODscreen is suitable for telephone interview and self-rating and may be useful in screening first-degree relatives, mixed populations and supposedly also the general public, but is not very useful in highly symptomatic, help-seeking individuals. Another screening instrument, the Prodromal Questionnaire (PQ), has items based on SIPS and the Schizotypal Personality Questionnaire (SPQ). This instrument is not sensitive to the threshold between prodromal and psychotic state. The Structured Interview of Prodromal Syndromes Screen (SIPS screen) instrument
Early detection

consists of 12 items covering positive symptoms only, the Youth Psychosis At Risk Questionnaire (Y-PARQ) is a self-report screening instrument based on the CAARMS for adolescents, and the Community Assessment of Psychic Experiences (CAPE) measures lifetime frequency of positive, negative and depressive symptoms and the level of distress associated with these symptoms.\textsuperscript{2,3}

The quality of assessment tools can be measured in various ways. ‘Reliability’ refers to the reproducibility of an instrument’s results across different assessors, settings and times. ‘Construct validity’ is the extent to which an instrument measures the theoretical construct it was designed to measure. This involves ‘convergent validity’, which is the degree of correlation between different scales measuring the same construct, confirming they are measuring the same thing; and ‘divergent validity’, which is the lack of correlation between scales measuring different constructs, confirming that they are measuring different things. Similarly, ‘known groups’ validity is the extent to which an instrument can demonstrate different scores for groups known to vary on the variables being measured. ‘Content validity’ is the extent to which each individual item on a scale represents the construct being measured, and ‘internal consistency’ is the degree of correlation between individual items within a scale.

‘Predictive validity’ refers to sensitivity, which is the proportion of correctly identified positives, and specificity, which is the proportion of correctly identified negatives. Sensitivity and specificity are measured by comparing an instrument’s results with known ‘gold standard’ results. ‘Responsiveness’ is the extent to which an instrument can detect clinically significant or practically important changes over time, and ‘area under the curve’ (AUC) is a global measure of test performance.

Method

We have included only systematic reviews (systematic literature search, detailed methodology with inclusion/exclusion criteria) published in full text, in English, from the year 2000 that report results separately for people with a diagnosis of schizophrenia, schizoaffective disorder, schizophreniform disorder or first episode schizophrenia. Due to the high volume of systematic reviews we have now limited inclusion to systematic meta-analyses. Where no systematic meta-analysis exists for a topic, systematic reviews without meta-analysis are included for that topic. As part of a wider search for all topics included in the library, reviews on early detection for schizophrenia were identified by searching the databases MEDLINE, EMBASE, CINAHL, Current Contents, PsycINFO and the Cochrane library. Hand searching reference lists of identified reviews was also conducted. When multiple copies of reviews were found, only the most recent version was included. The decision to include or exclude reviews was conducted in duplicate by two reviewers with any disagreements settled by discussion. All quality assessments and data extraction have been completed in duplicate by two independent reviewers who were not masked to review authors.

Review reporting assessment was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist that describes a preferred way to present a meta-analysis\textsuperscript{4}. Reviews with less than 50% of items checked have been excluded from the library. The PRISMA flow diagram is a suggested way of providing information about studies included and excluded with reasons for exclusion. Where no flow diagram has been presented by individual reviews, but identified studies have been described in the text, reviews have been
Early detection

checked for this item. Note that early reviews may have been guided by less stringent reporting checklists than the PRISMA, and that some reviews may have been limited by journal guidelines.

Evidence was graded using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group approach where high quality evidence such as that gained from randomised controlled trials (RCTs) may be downgraded to moderate or low if review and study quality is limited, if there is inconsistency in results, indirect comparisons, imprecise or sparse data and high probability of reporting bias. It may also be downgraded if risks associated with the intervention or other matter under review are high. Conversely, low quality evidence such as that gained from observational studies may be upgraded if effect sizes are large, there is a dose dependent response or if results are reasonably consistent, precise and direct with low associated risks (see end of table for an explanation of these terms). The resulting table represents an objective summary of the available evidence, although the conclusions are solely the opinion of staff of NeuRA (Neuroscience Research Australia).

Results

We found seven systematic reviews that met our inclusion criteria. These are presented below the results in alphabetical order.

• Moderate to high quality evidence suggests excellent sensitivity and modest specificity for psychometric tools that assess the risk of developing psychosis in help-seeking individuals referred to high-risk services.

• Moderate to high quality evidence indicates the mean rate of transition to full psychotic episode in clinical high risk groups is around 29% after up to 3 years following assessment of risk. Studies with older samples reported higher transition rates than studies with younger samples, and more recent publications reported lower transition rates than older publications. Studies using the basic symptoms approach reported higher transition rates than studies using the ultra-high risk approach. Studies of people receiving psychosocial treatments (e.g. cognitive behavioral therapy) reported lower transition rates than studies of people receiving standard care (e.g. case management). Studies of people on antipsychotics reported lower transition rates than studies of people not on antipsychotics.

• Moderate quality evidence suggests an increased rate of a diagnosis of schizophrenia compared to affective psychosis at ~2.5 year follow-up in people who were previously assessed as being at high risk for psychosis. This risk is highest for older people and those assessed for risk using the basic symptoms criteria.

• Moderate quality evidence suggests the CAPE has good internal reliability, representing the same construct, with three factors: positive, negative and depressive symptoms.

• Moderate to low quality evidence suggests the pretest risk for psychosis in help-seeking people is around 15%, with highest pretest risk in studies recruiting primarily from mental health services.
**Chuma J, Mahadun P**

*Predicting the development of schizophrenia in high-risk populations: systematic review of the predictive validity of prodromal criteria*

*The British Journal of Psychiatry 2011; 199: 361-366*

View review abstract online

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity and specificity of instruments that assess high-risk mental states.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of evidence</td>
<td>Moderate to high quality evidence (large samples, consistent, direct, appears precise) suggests instruments based on ultra-high risk criteria have moderate sensitivity and specificity. Moderate to low quality evidence (1 small to medium-sized study) suggests the BSABS, which is based on the basic symptoms approach, has good sensitivity and moderate specificity.</td>
</tr>
</tbody>
</table>

**Ultra-high risk criteria**

*Authors conclude that the ultra-high risk criteria has moderate sensitivity and specificity;*

12 studies, N = 1918, mean follow-up period = 18.6 months

- Sensitivity = 0.66, 95%CI 0.61 to 0.70
- Specificity = 0.73, 95%CI 0.71 to 0.75

*Excluding two studies (outliers) that did not systematically enrol participants increased sensitivity but not specificity;*

- 10 studies, N = 1444
- Sensitivity = 0.81, 95%CI 0.76 to 0.85
- Specificity = 0.67, 95%CI 0.64 to 0.70

**Basic symptoms criteria**

*Measured using the Bonn Scale for the Assessment of Basic Symptoms (BSABS)*

*Authors conclude that basic symptoms criteria has good sensitivity and moderate specificity;*

1 study, N = 160, follow-up period = 9.6 years

- Sensitivity = 0.97, 95% CI 0.91 to 1.00
- Specificity = 0.59, 95% CI 0.48 to 0.70

**Consistency in results**

*Authors report the ultra-high risk approach studies’ results were consistent.*
Early detection

<table>
<thead>
<tr>
<th>Precision in results</th>
<th>Appears precise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directness of results</td>
<td>Direct</td>
</tr>
</tbody>
</table>


The Dark Side of the Moon: Meta-analytical Impact of Recruitment Strategies on Risk Enrichment in the Clinical High Risk State for Psychosis


Comparison

Pretest risk of psychosis; the probability of developing psychosis before a test result is known which depends on the underlying risk in the population being tested.

Summary of evidence

Moderate to low quality evidence (large sample, inconsistent, appears imprecise, direct) suggests the pretest risk for psychosis in help-seeking people is around 15% with highest risk in studies recruiting primarily from mental health services.

Study recruitment techniques

11 studies, N = 2519, mean follow-up period = 38 months

The pretest risk for psychosis in help-seeking patients was around one-sixth of the sample; 15%, 95%CI 9% to 24%, I² = 96%, p < 0.001

Studies directing their outreach campaigns to mental health services had higher pretest risk of psychosis than those reaching out to the general public and those with a high proportion of self-referrals. Studies with intensive outreach campaigns showed reduced pretest risk of psychosis.

Authors report no evidence of publication bias.

<table>
<thead>
<tr>
<th>Consistency in results</th>
<th>Inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision in results</td>
<td>Appears imprecise</td>
</tr>
<tr>
<td>Directness of results</td>
<td>Direct</td>
</tr>
</tbody>
</table>

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au
To donate, phone 1800 888 019 or visit www.neura.edu.au/donate/schizophrenia
**At risk or not at risk? A meta-analysis of the prognostic accuracy of psychometric interviews for psychosis prediction**


**World Psychiatry 2015; 14(3): 322-32**

[View review abstract online](#)

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Predictive validity of psychometric tools that assess the risk of developing psychosis in help-seeking individuals referred to high-risk services. Mean follow-up was 38 months.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of evidence</td>
<td>Moderate to high quality evidence (large sample, inconsistent, appears precise, direct) suggests excellent sensitivity and modest specificity for psychometric tools that assess the risk of developing psychosis in help-seeking individuals referred to high-risk services.</td>
</tr>
</tbody>
</table>

**Predictive validity for transition to psychosis**

**High-risk assessment tools:** CAARMS, SIPS, BSIP, BSABS, SPI-A, SPI-CY

**Psychosis outcome assessment tools:** ICD, DSM, BPRS or CAARMS

*Excellent sensitivity and moderate specific was reported. Testing positive for clinical high risk was associated with a 26% risk of developing psychosis within 38 months, and testing negative for clinical high risk was associated with a 1.56% risk of developing psychosis;*

- Sensitivity: 11 studies, N = 2519, 0.96, 95%CI 0.92 to 0.98, I² 55%, p = 0.02
- Specificity: 11 studies, N = 2519, 0.47, 95%CI 0.38 to 0.57, I² 95%, p = 0.001

Area Under the Curve: 11 studies, N = 2519, AUC 0.90, 95%CI 0.87 to 0.93

Meta-regression analyses revealed a moderating effect for exposure to antipsychotics, such that there was significantly less sensitivity in the five studies where subjects were exposed to antipsychotics than in the six studies where subjects were not exposed to antipsychotics (0.94 vs. 0.98). No moderating effects were found for age, sex, follow-up time, sample size, study quality, or proportion of clinical high-risk individuals.

Authors report low positive predictive value in; the general population, unselected psychiatric adolescent samples, patients accessing public treatment or primary care services, patients admitted to forensic units, post-partum women, ethnic minorities, military, refugees, patients with epilepsy, and prisoners.
Early detection

<table>
<thead>
<tr>
<th>Consistency in results</th>
<th>Inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision in results</td>
<td>Appears precise</td>
</tr>
<tr>
<td>Directness of results</td>
<td>Direct</td>
</tr>
</tbody>
</table>

Fusar-Poli P, Bechdolf A, Taylor M, Bonoldi I, Carpenter W, Yung A, McGuire P

At Risk for Schizophrenic or Affective Psychoses? A Meta-Analysis of DSM/ICD Diagnostic Outcomes in Individuals at High Clinical Risk


View review abstract online

Comparison

Diagnostic outcomes of people who were assessed as being at high risk for psychosis.

Summary of evidence

Moderate quality evidence (large sample, inconsistent, imprecise, direct) suggests a higher risk of a diagnosis of schizophrenia than affective psychosis in people who transitioned to psychosis, particularly older people and people assessed using the basic symptoms approach.

Diagnosis

23 studies, N = 2182, mean follow-up period = 2.35 years

Around one third (26%) of the overall sample transitioned to psychosis.

Males were more likely to transition to psychosis than females;

RR 1.82, 95% CI 1.33 to 2.49, p < 0.001

More people who transitioned to psychosis were given a diagnosis of schizophrenia than a diagnosis of affective psychosis;

RR 5.43, 95% CI 3.35 to 8.83, p < 0.001, I² 42%, p < 0.001

This effect was significantly more pronounced in people assessed for risk using the basic symptoms criteria than using the ultra-high risk criteria;

Basic symptoms criteria: RR = 17.07

Ultra-high risk criteria: RR = 3.81

Q = 21.108, p < 0.001

This effect increased significantly with increased mean study age;
Early detection

$b = 0.14$, 95% CI 0.04 to 0.23, $p = 0.005$

There were no moderating effects according to diagnostic outcome tool (ICD vs. DSM), medication (treated vs. untreated), publication year, duration of follow-up, or study quality.

Authors report no evidence of publication bias.

<table>
<thead>
<tr>
<th>Consistency in results</th>
<th>Inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision in results</td>
<td>Imprecise</td>
</tr>
<tr>
<td>Directness of results</td>
<td>Direct</td>
</tr>
</tbody>
</table>


Predicting psychosis: meta-analysis of transition outcomes in individuals at high clinical risk

Archives of General Psychiatry 2012; 69(3): 220-229

View review abstract online

Comparison

Rates of transition to psychosis in people at high clinical risk due to having attenuated or prodromal symptoms of psychosis.

Summary of evidence

Moderate to high quality evidence (large samples, inconsistent, appears precise, direct) indicates the mean risk of transition to full psychotic episode in clinical high risk groups is around 29.2% over 31 months, with transition rate increasing over time.

Studies with older samples reported higher transition rates than studies with younger samples, and more recent publications reported lower transition rates than older publications. Studies using the basic symptoms approach reported higher transition rates compared to studies using the ultra-high risk approach.

Studies of people receiving psychosocial treatments (e.g. cognitive behavioral therapy) reported lower transition rates than studies of people receiving standard care (e.g. case management), and studies of people receiving antipsychotics reported lower transition rates than studies of people not receiving antipsychotics.

No differences were reported between studies defining “high risk” or “transition to psychosis” using the Structured Interview

To donate, phone 1800 888 019 or visit www.neura.edu.au/donate/schizophrenia
Early detection for Prodromal Syndromes or the Comprehensive Assessment of At-Risk Mental States.

<table>
<thead>
<tr>
<th>Rates of transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>27 studies, N = 2502, mean follow-up period = 31 months</td>
</tr>
<tr>
<td>Transition to psychosis occurred in about one third of people identified as being at risk;</td>
</tr>
<tr>
<td>Transition rate: 29.2%, 95%CI 27.3% to 31.1%, $I^2 = 83.11%$, $p &lt; 0.001$</td>
</tr>
<tr>
<td>Removing low quality studies resulted in a decrease in the overall estimate to 22%.</td>
</tr>
<tr>
<td>The risk of transition to psychosis increased with time;</td>
</tr>
<tr>
<td>6 months after initial presentation, transition rate: 18%</td>
</tr>
<tr>
<td>1 year after initial presentation, transition rate: 22%</td>
</tr>
<tr>
<td>2 years after initial presentation, transition rate: 29%</td>
</tr>
<tr>
<td>3 years after initial presentation, transition rate: 36%</td>
</tr>
<tr>
<td>Increasing mean age of study participants was related to increased rate of transition;</td>
</tr>
<tr>
<td>$\beta = 0.07$, 95%CI 0.05 to 0.09, $p &lt; 0.001$</td>
</tr>
<tr>
<td>Studies using the basic Symptoms approach reported significantly higher transition rates than studies using the Ultra-high risk approach;</td>
</tr>
<tr>
<td>Basic symptoms: 2 studies, 48.5%, 95%CI 41.9% to 55.9%</td>
</tr>
<tr>
<td>Ultra-high Risk: 22 studies, 27.7%, 95%CI 25.6% to 29.9%</td>
</tr>
<tr>
<td>$Q_B = 46.56$, $p &lt; 0.001$</td>
</tr>
<tr>
<td>There were no significant differences in transition rates between studies using the Structured Interview for Prodromal Syndromes (SIPS) or the Comprehensive Assessment of At-Risk Mental States (CAARMS).</td>
</tr>
<tr>
<td>Studies using ICD-10, DSM-III, or DSM-IV diagnostic criteria for psychosis reported a transition rate of around half of the study samples;</td>
</tr>
<tr>
<td>51.1%, 95%CI 43.4% to 58.7%</td>
</tr>
<tr>
<td>More recent publications reported lower transition rates;</td>
</tr>
<tr>
<td>$\beta = -0.15$, 95%CI -0.17 to -0.11, $p &lt; 0.001$</td>
</tr>
<tr>
<td>Studies of people receiving psychosocial treatments (e.g. cognitive behavioural therapy) reported</td>
</tr>
</tbody>
</table>
Early detection

lower transition rates than studies of people receiving standard care (e.g. case management);
Cognitive behavioural therapy: 24.9%, 95%CI, 23.2% to 28.0%
Standard care: 32.8%, 95% CI, 29.5% to 36.2%
\( Q_B = 11.69, \ p < 0.001 \)

Studies of people on antipsychotics reported lower mean transition risk than studies of people not on antipsychotics;
Antipsychotics: 22.9%, 95% CI 20.5% to 25.5%
No antipsychotics: 36.5%, 32.1% to 41.3%
\( Q_B = 28.32, \ p < 0.001 \)

There was no effect of sex on rates of transition;
\( \beta = 0.002, \ 95\%CI = 0.08 \text{ to } 0.12, \ p = 0.88 \)

<table>
<thead>
<tr>
<th>Consistency in results</th>
<th>Inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision in results</td>
<td>Appears precise</td>
</tr>
<tr>
<td>Directness of results</td>
<td>Direct</td>
</tr>
</tbody>
</table>

**Mark W, Toulopoulou T**

**Psychometric Properties of "Community Assessment of Psychic Experiences": Review and Meta-analyses**

Schizophrenia Bulletin 2016; 42(1): 34-44

View review abstract online

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Exploratory factor analysis and internal reliability of the Community Assessment of Psychic Experiences (CAPE).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of evidence</td>
<td>Moderate quality evidence (large sample, unable to assess consistency or precision, direct) suggests the CAPE has good internal reliability, with 3 factors; positive, negative and depressive symptoms.</td>
</tr>
</tbody>
</table>

Exploratory factor analysis; identifies the underlying relationships between measured variables

Internal reliability; the degree to which items on the scale measure the same construct
Factor analysis confirmed a 3-factor model consisting of positive, negative, and depressive subscales, which accounted for 93.1% of the variance in scores;

The Positive dimension consisted of “bizarre experiences”, “delusional ideations”, and “perceptual anomalies”.

The Negative dimension consisted of “social withdrawal”, “affective flattening”, and “avolition/lack of motivation”.

CAPE showed good internal reliability;
18 samples, N ~ 77,191
CAPE-42 full scale reliability mean = 0.91
CAPE-positive subscale reliability mean = 0.84
CAPE-negative subscale reliability mean = 0.81
CAPE-depressive subscale reliability mean = 0.76

Subgroup analysis of sample age revealed CAPE-positive and CAPE-negative subscales had greater internal reliability in younger samples (≤ 25 years old) than in older samples (> 25 years old), with no differences on the full scale CAPE or CAPE-depressive subscale.

<table>
<thead>
<tr>
<th>Consistency in results</th>
<th>No measure of heterogeneity is provided.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision in results</td>
<td>No confidence intervals are provided.</td>
</tr>
<tr>
<td>Directness of results</td>
<td>Direct</td>
</tr>
</tbody>
</table>

Olsen KA, Rosenbaum B

Prospective investigations of the prodromal state of schizophrenia: assessment instruments

Acta Psychiatrica Scandinavica 2006; 113(4): 273-82
View review abstract online

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Description of instruments for screening and assessing at risk mental states.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening</td>
<td>PRODscreen, PQ, Y-PARQ and SIPS screen</td>
</tr>
<tr>
<td>Assessing at risk mental state</td>
<td>- Triad of mental states</td>
</tr>
<tr>
<td></td>
<td>CAARMS and SIPS</td>
</tr>
</tbody>
</table>
Early detection

- **Basic symptoms approach (BS)**
  BSABS, SPI-A,
  - **Multiple factors**
  EASE and ERiraos

**Summary of evidence**

Unable to fully assess the quality of evidence – no diagnostic meta-analysis, and limited data.

Authors conclusions;

*For screening:* All instruments are still under validation

*For assessing at risk mental state:* The psychometric properties of the instruments are only at a preliminary stage of investigation. The results are based on highly selected clinical populations, and cannot be generalized to other populations

- **BS approach**
  The presence of BS (measured by BSABS) has shown the most significant predictive values of later schizophrenia, but on a very long follow up interval (average 9.6 years) only

- **APS approach**
  Both CAARMS and SIPS seem valid in detecting samples at increased risk of psychosis although only a minority develop full-blown psychosis

**Screening**

**Reliability and Validity**

**PRODscreen**

1 study (N = 132 mixed sample of research subjects) correctly identified a SIPS-defined at-risk mental state in 77% of the sample, showing good concurrent validity.

Sensitivity is 80% and specificity is 75%.

**PQ**

1 study (N = 113 subjects referred to an early detection and intervention clinic) - good concurrent validity of PQ-positive subscale against SIPS.

With a cut-off at 8 or more items of positive symptoms - sensitivity is 90%, specificity 49%, and with a cut-off at 14 or more items of positive symptoms - sensitivity is 71% and specificity is 81%.

**Y-PARQ**

1 study (N = 74 adolescents potentially displaying at risk mental states) - PPV of a CAARMS defined at-risk mental state = 82.4%, showing good concurrent validity.

**SIPS screen**

1 study (N = 36 sample of subjects referred for evaluation of at risk mental state). Sensitivity is 90% and 100% is specificity
### Early detection

#### Assessing a possible early prodromal state - Basic symptoms approach (BS)

**Reliability and Validity**

BSABS

1 study (N = 110 out-patients with at least one BS) - 70% developed schizophrenia in an average follow-up of 9.6 years (specificity 0.59, PPV 80%). In the control group of 50 outpatients without BS, the absence of BS excluded schizophrenia with a probability of 96% (sensitivity 0.98, NPV 98.7%).

Ten symptoms predicted schizophrenia with a probability of 71–91%. Reanalysis if these data suggests two of the 10 items – thought pressure and decreased ability to discriminate between ideas/perception and phantasy/true memory – were diagnostically relevant.

SPI-A

1 study (N = 147) 17% who have reported experiencing at least one BS have developed schizophrenia within an average of 12 months (PPV 17%).

Preliminary results indicate good inter-rater reliability (89%) and ‘good’ construct validity.

#### Assessing a possible late prodromal state - attenuated positive symptoms approach (APS)

**Reliability and Validity**

CAARMS

1 study (N = 150 non-psychotic, help-seeking individuals), 6 month follow-up; sensitivity = 83%, specificity = 74%, PPV = 12%, NPV = 99%.

Inter-rater reliability (N = 34, UHR) range of 0.62–0.93

SIPS

1 study (N= 13 with a SIPS-defined prodromal state), 46% developed psychosis within 6 months, 54% within 12 months. Agreement on prodromal/non-prodromal status of 18 subjects = 93%.

Another larger study (N = 34, treatment seeking sample) has reported preliminary data on transition rates at 6, 12, 18 and 24 months: 43%, 50%, 62% and 67% respectively (sensitivity 100%, specificity 73% at 24 months).

#### Assessing a possible at risk mental state – other instruments
Early detection

ERIraos
This instrument has not yet been published, and is currently being validated.

EASE
This scale was not specifically developed for the assessment of at-risk mental state, however it explores phenomenological aspects in the pre-onset phase focusing on experiential anomalies of self-awareness and disorders in the subjective experience.

Reliability and Validity
EASE
Inter-rater reliability – 1 study (N = 14 in-patients) - Kappa reliability 0.6 to 1.0. Test–retest reliability is under evaluation.

<table>
<thead>
<tr>
<th>Consistency in results</th>
<th>Precision in results</th>
<th>Directness of results</th>
</tr>
</thead>
<tbody>
<tr>
<td>No measure of heterogeneity is provided.</td>
<td>No confidence intervals are provided.</td>
<td>Direct</td>
</tr>
</tbody>
</table>

Explanation of acronyms

Received probability, BS = Basic Symptoms, BSABS = Bonn Scale for the Assessment of Basic Symptoms, BSIP = Basel Screening Instrument for Psychosis, CAARMS = Comprehensive Assessment of At-Risk Mental States, CAPE = Community Assessment of Psychic Experiences, CI = Confidence Interval, DSM = Diagnostic and Statistical Manual of Mental Disorders, ERIraos = Early Recognition Inventory EASE = Examination of Anomalies in Self-experience, ICD = International Classification of Diseases, I² = the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance), N = number of participants, NPV = negative predictive value - the proportion of patients with negative test results who are correctly diagnose, PANSS = Positive and negative syndrome scale, p = statistical probability of obtaining that result (p < 0.05 generally regarded as significant), PPV = positive predictive value - proportion of patients with positive test results who are correctly diagnosed, PQ = Prodomal Questionnaire, PRODscreen = Prodromal screening test, Q = Q statistic for the test of heterogeneity, Qw = test for within group differences (heterogeneity in study results within a group of studies – measure of study consistency), Qb = test for between group differences (heterogeneity between groups of studies for an outcome of interest), RR = relative risk, SIPS = Structured Interview of Prodromal Syndromes, SPI-A = Schizophrenia Prediction Instrument – Adult version, UHR = Ultra High Risk for psychosis, Y-PARQ = Youth Psychosis At Risk Questionnaire, vs. = versus
Explanation of technical terms

*Bias has the potential to affect reviews of both RCT and observational studies. Forms of bias include; reporting bias – selective reporting of results, publication bias - trials that are not formally published tend to show less effect than published trials, further if there are statistically significant differences between groups in a trial, these trial results tend to get published before those of trials without significant differences; language bias – only including English language reports; funding bias - source of funding for the primary research with selective reporting of results within primary studies; outcome variable selection bias; database bias - including reports from some databases and not others; citation bias - preferential citation of authors. Trials can also be subject to bias when evaluators are not blind to treatment condition and selection bias of participants if trial samples are small.†

† Different effect measures are reported by different reviews.

Prevalence refers to how many existing cases there are at a particular point in time. Incidence refers to how many new cases there are per population in a specified time period. Incidence is usually reported as the number of new cases per 100,000 people per year. Alternatively some studies present the number of new cases that have accumulated over several years against a person-years denominator. This denominator is the sum of individual units of time that the persons in the population are at risk of becoming a case. It takes into account the size of the underlying population sample and its age structure over the duration of observation.

Reliability and validity refers to how accurate the instrument is. Sensitivity is the proportion of actual positives that are correctly identified (100% sensitivity = correct identification of all actual positives) and specificity is the proportion of negatives that are correctly identified (100% specificity = not identifying anyone as positive if they are truly not).

Weighted mean difference scores refer to mean differences between treatment and comparison groups after treatment (or occasionally pre to post treatment) and in a randomized trial there is an assumption that both groups are comparable on this measure prior to treatment. Standardized mean differences are divided by the pooled standard deviation (or the standard deviation of one group when groups are homogenous) that allows results from different scales to be combined and compared. Each study’s mean difference is then given a weighting depending on the size of the sample and the variability in the data. 0.2 represents a small effect, 0.5 a moderate effect, and 0.8 and over represents a large treatment effect†.

Odds ratio (OR) or relative risk (RR) refers to the probability of a reduction (< 1) or an increase (> 1) in a particular outcome in a treatment group, or a group exposed to a risk factor, relative to the comparison group. For example, a RR of 0.75 translates to a reduction in risk of an outcome of 25% relative to those not receiving the treatment or not exposed to the risk factor. Conversely, an RR of 1.25 translates to an increased risk of 25% relative to those not receiving treatment or not having been exposed to a risk factor. An RR or OR of 1.00 means there is no difference between groups. A medium effect is considered if RR > 2 or < 0.5 and a large effect if RR > 5 or < 0.2†. InOR stands for logarithmic OR where a lnOR of 0 shows no difference between groups. Hazard ratios
Early detection

measure the effect of an explanatory variable on the hazard or risk of an event.

Correlation coefficients (eg, r) indicate the strength of association or relationship between variables. They are an indication of prediction, but do not confirm causality due to possible and often unforeseen confounding variables. An r of 0.10 represents a weak association, 0.25 a medium association and 0.40 and over represents a strong association. Unstandardised (b) regression coefficients indicate the average change in the dependent variable associated with a 1 unit change in the dependent variable, statistically controlling for the other independent variables. Standardised regression coefficients represent the change being in units of standard deviations to allow comparison across different scales.

‡ Inconsistency refers to differing estimates of treatment effect across studies (i.e. heterogeneity or variability in results) that is not explained by subgroup analyses and therefore reduces confidence in the effect estimate. I² is the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance) - 0% to 40%: heterogeneity might not be important, 30% to 60%: may represent moderate heterogeneity, 50% to 90%: may represent substantial heterogeneity and 75% to 100%: considerable heterogeneity. I² can be calculated from Q (chi-square) for the test of heterogeneity with the following formula:

\[ I^2 = \left( \frac{Q - df}{Q} \right) \times 100\% \]

§ Imprecision refers to wide confidence intervals indicating a lack of confidence in the effect estimate. Based on GRADE recommendations, a result for continuous data is considered imprecise if the upper or lower confidence limit crosses an effect size of 0.5 in either direction, and for binary and correlation data, an effect size of 0.25. GRADE also recommends downgrading the evidence when sample size is smaller than 300 (for binary data) and 400 (for continuous data), although for some topics, this criteria should be relaxed\(^1\).
Early detection

References