

Review article

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

Jeffer Chuma and Prem Mahadun

Background

There is a great deal of debate on the usefulness and accuracy of prodromal criteria in predicting schizophrenia. The risk of treating people who screen false positive with medication is considerable. Yet intervening during the prodromal stage of illness may reduce the burden caused by schizophrenia.

Aims

To draw together the evidence base for the predictive validity of prodromal criteria in identifying individuals at high risk of developing schizophrenia.

Method

We conducted a systematic review of prospective studies investigating the predictive validity of prodromal criteria in schizophrenia.

Results

Our study found two main criteria, ultra-high-risk criteria and basic-symptoms criteria, used in studies investigating the predictive validity of prodromal symptoms. The sensitivity and specificity of ultra-high-risk criteria was 0.81 (95% CI 0.76–0.85) and 0.67 (95% CI 0.64–0.70) respectively and for basic-symptoms criteria sensitivity and specificity was 0.97 (95% CI 0.91–1.00) and 0.59 (95% CI 0.48–0.70) respectively.

Conclusions

Both ultra-high-risk criteria and basic-symptoms criteria are useful in predicting the development of schizophrenia among high-risk populations.

Declaration of interest

None.

This review seeks to contribute to the important debate on the usefulness of prodromal symptoms in identifying people at high risk of developing schizophrenia. Research suggests that schizophrenia has a prodromal period where significant changes from premobid functioning can be observed.¹ Prodromal symptoms of schizophrenia are defined as early specific and non-specific symptoms preceding the first psychotic episode.^{2–4} They consist of behavioural abnormalities, pseudoneurotic signs and subtle cognitive and affective changes.⁴ In general, it is believed that individuals who will go on to develop schizophrenia go through a variety of abnormal, subjective experiences that progressively develop during pre-puberty and puberty.² However, an individual could present with the same prodromal symptoms without necessarily developing a psychotic illness.⁴

There are generally two viewpoints pertaining to the subject of prodromal symptoms,⁵ the basic-symptoms⁶ approach and the ultra-high-risk approach.^{2,7} The basic-symptoms approach refers to the validity of particular neuropsychological symptoms, describing disturbances prior to the onset of psychosis.⁵ These disturbances are assessed in the domains of perception, cognition, motor functioning, will, energy level and stress tolerance.⁶ The ultra-high-risk criteria are composed of the following three alternative risk syndromes: attenuated positive symptoms, brief limited intermittent psychotic symptoms (BLIPS) and trait and state risk factors.² Attenuated positive symptoms is used when the individual has symptoms that deviate from normal phenomena but that are not yet frankly psychotic.² Brief limited intermittent psychotic symptoms is used when the symptoms are of psychotic intensity but they are very infrequent, or have a total duration of less than 7 days before resolving spontaneously.² Whereas the term trait and state risk factors is used when there are non-specific symptoms such as lowered mood or anxiety symptoms plus some trait risk-factors for psychotic disorder, either schizotypal personality disorder or a family history of psychotic disorder in a first-degree relative.² With both

approaches (basic symptoms and ultra-high risk), at this stage, neither the recognition nor the description of the period preceding psychosis allows for the effective prediction of psychosis.³

Consideration of the predictive validity of prodromal symptoms is important for both researchers and clinicians; it evokes a lot of debate on the usefulness and accuracy of prodromal screening instruments given the ethical implications that comes with it.^{8–10} The risk of negative effects for the large numbers of people who screen false positive (people who do not convert to schizophrenia) are considerable. The rate of false positives in some studies has been as high as 60–90%¹¹ leading to unfavourable risk to benefit ratios in studies with lower conversion rates. These risks include unnecessary fear of illness, restriction of life goals, use of medication and their side-effects, stigmatisation and loss of confidentiality and insurability.^{11,12} However, early identification and intervention during the prodromal phase of the illness promises to reduce the economic burden caused by schizophrenia. Ruhrmann *et al* further contributed to this debate by suggesting that the majority of help-seeking at-risk people fulfil DSM-IV general criteria for mental disorders and have a need for and right to treatment.¹³ They argue for prodromal symptoms to be classified as psychosis-spectrum disorder in DSM-5.¹³

Method**Search method**

A comprehensive search of electronic databases for diagnostic research reports was carried out. The following electronic health-related databases were comprehensively searched, CINAHL (1982–2010), EMBASE (1980–2010) MEDLINE (1950–2010) and PsycINFO (1806–2010). The search terms were identified using the population, intervention, comparison and outcome method.¹⁴ We used the following MeSH terms: schizophrenia, psychosis

disorders and prodrome, together with the text terms: prodromal symptoms, prodromal scales, prodromal\$, early-intervention and screening. These were used in combination with the search terms recommended for diagnostic reviews:¹⁵ sensitivity, negative-predictive-values, positive-predictive-values, likelihood-ratios, ROC curves and diagnostic odds ratios. Reference lists of articles obtained were checked to identify relevant reports. This was done in a systematic way following up references from one article to another to identify possible studies.

Selection criteria

We included studies with prospective or cohort designs and cross-sectional designs. The primary studies had a clearly specified population, from which a prodromal criterion was administered to identify clearly those with prodromal symptoms from those without. The two groups were then followed up for a number of months and assessed again with a diagnostic instrument to determine those who had converted to schizophrenia. The exclusion criteria were the presence of psychotic disorder (as we were interested in assessing those at risk of psychotic disorder), known organic cause of presentation and known intellectual disability. The gold standard (reference standard) for this review was DSM-IV.¹⁶

Validity assessment

Data were independently extracted by two researchers. We took care to avoid the 'double counting' of evidence, particularly where the same first authors were quoted in several validation studies. We assessed study quality in line with accepted guidelines.^{17,18} In particular we sought information on the application of a diagnostic standard independent of the knowledge of scores on the prodromal criteria ('masked'). Lack of masking is a potential source of bias within cross-sectional validation studies, since foreknowledge of test scores by those applying a diagnostic gold standard can create an exaggerated level of agreement.¹⁹ We paid particular attention to the description of the study population and of the diagnostic tests, since bias in estimation of test accuracy has been associated with poor description of these.²⁰ The use of the appraisal tools helps in guarding against different forms of bias, for example, inappropriate use of the reference standard, lack of masking and a case-control study design.²⁰ Where possible raw data rather than percentages or measures of effect were extracted from primary studies; this helps in avoiding hidden errors in data extraction. In the case of missing data during the data extraction process, the original authors of the primary studies were contacted for clarification. Data requiring manipulation were first extracted in their original form and then transformed in subsequent steps.^{15,20}

Quantitative data synthesis

We first constructed 2 × 2 tables for all studies. From these we calculated sensitivity, specificity, likelihood ratios (positive and negative), diagnostic odds ratio (OR) and a summary receiver operator characteristic (ROC) curve using Meta-Disc software version 1.4 for Windows (www.hrc.es/investigacion/metadisc_en.htm).²¹ The likelihood ratio represents a measure of the predictive ability of a test that, unlike positive predictive value, is a fundamental predictive attribute of the instrument, which does not vary according to the baseline prevalence of the disorder in question.²⁰ The diagnostic OR is the ratio of the odds of a positive test among those with the disorder to the odds of a positive result among those without the disorder.^{15,20} This is the recommended metric in diagnostic meta-analyses.²⁰ Receiver operator characteristic curves are the most informative way of representing the inherent trade-off between sensitivity and

specificity for a test of a diagnostic instrument.²² We therefore created a single plot of sensitivity and specificity in ROC space summarising each study, weighted by study size. Summary ROC²² curves were then constructed using a bivariate model to produce a 95% confidence ellipse within ROC space.²⁰ Unlike a traditional ROC plot that explores the effect of a varying threshold on sensitivity and specificity in a single study, each data point in the summary ROC space represents a separate study.²² Between-study heterogeneity was assessed using the I^2 statistic of the pooled diagnostic OR,²³ which describes the percentage of total variation across studies that is caused by heterogeneity rather than chance. The I^2 statistic has several advantages over other measures of heterogeneity (such as chi-squared), including greater statistical power to detect clinical heterogeneity when fewer studies are available.²³ As a guide, I^2 values of 25% may be considered 'low', 50% 'moderate' and 75% 'high'. Where there was significant between-study heterogeneity, we sought to explore the causes of this heterogeneity.²³

Results

Literature search

Our searches identified 584 potential studies. The titles of the 584 were read by two reviewers and 417 of the citations were excluded at this stage as they did not meet our inclusion criteria. The titles and abstracts for the remaining 167 articles were read by one reviewer (J.C.). Out of these, 127 studies were excluded with reasons ranging from the articles not being diagnostic studies to articles investigating reliability of instruments rather than the diagnosis accuracy of the criteria. Of the 40 studies selected from examination of the complete text, 24 studies were excluded because they did not meet the criteria in the following areas: index test, gold standard, study design aspects and outcome type. Of these, two were excluded because of multiple publications. Two papers by Miller *et al* reported the same study in different journals under different titles; we included one paper³ instead of both. Similarly, a study by Klosterkötter *et al* was published twice and we included only one report²⁴ instead of both.

Methodological quality of studies

All studies included in this review were assessed against the standard for reporting of diagnostic accuracy (STARD)¹⁷ checklists. One investigator assessed the quality of all included studies. To evaluate the inter-observer variability in the rating of the STARD criteria, a second investigator examined three randomly selected publications, masked to the results of the first investigator. Six studies^{25–30} did not describe in detail the number, training, and expertise of the persons executing and reading the index tests and the reference standard (STARD item 10). All studies^{3,24–37} did not specify whether investigators were masked to the results of the other test (index test or reference test; item 11 of the STARD checklist). One of the main requirements for the assessment of the predictive validity of diagnostic studies is the inclusion of participants not meeting prodromal criteria yet similarly recruited as those fulfilling the criteria to avoid spectrum bias¹⁷ (STARD item 16). Three studies^{31,34,37} did not systematically enrol participants who tested negative to the prodromal criteria. We identified only three studies^{24,29,33} with sufficient sample sizes and well-defined control groups similarly recruited to those fulfilling prodromal criteria.

Study characteristics

We found 13 prospective evaluation studies^{3,24–31,33–35,37} meeting our inclusion criteria and with sufficient data (reported in 15

publications,^{3,25–38} see online Table DS1). Twelve studies examined the predictive validity of the ultra-high-risk criteria,^{3,25–30,33–35} two studies examined the predictive validity of the basic-symptoms criteria.^{24,36} All included studies made a reference to a DSM diagnosis of schizophrenia, established according to standardised methods by a trained research worker or mental health professional. Three studies^{3,30,35} had very small sample sizes of 34, 30 and 34 participants respectively, however, with similarly recruited control groups to the high-risk group. All identified prospective studies^{3,25–38} were from specialised early detection assessment settings that suggest selection bias and therefore makes it difficult to generalise the predictive validity of these studies across different clinical setting.

Meta-analysis of ultra-high risk criteria for predicting schizophrenia

We pooled 12 studies (1918 participants: 402 confirmed with schizophrenia by DSM gold standard).^{3,25–31,33–35,37} When we combined psychometric attributes across studies, we found a low level of between-study heterogeneity²³ (combined diagnostic OR $I^2 = 35.5\%$). Pooled sensitivity was 0.66 (95% CI 0.61–0.70) and specificity was 0.73 (95% CI 0.71–0.75). The likelihood ratio for a positive test was 3.53 (95% CI 2.66–4.69), and the likelihood ratio for a negative test was 0.33 (95% CI 0.22–0.51). When we summarised individual studies within ROC space, we found that the majority of studies were gathered within an informative top left-hand corner with an area under curve (AUC) of 0.85 (Fig. 1). However, two studies were obvious ‘outliers’: a multisite longitudinal study predicting psychosis in youth at high risk³¹ and a prospective European prediction of psychosis study.³⁴ These studies each had relatively low sensitivity. The omission of these studies eliminated the level of between-study heterogeneity from combined diagnostic OR $I^2 = 35.5\%$ to $I^2 = 0.0\%$ as shown in Fig. 2. Heterogeneity was therefore explained by the recruitment methods used in the two studies;^{31,34} both studies did not systematically enrol participants who were ultra-high-risk criteria negative. The predictive values in these studies were calculated from adjunctive criteria, such as demographics, functioning and family history among the ultra-high-risk positive population. The overall pooled result was substantially altered by the exclusion of these studies. When these outliers were omitted the pooled sensitivity was 0.81 (95% CI 0.76–0.85) and specificity was 0.67 (95% CI 0.64–0.70). The average follow-up period for 11 of the included studies^{3,25–31,33–35} was 18.6 months and one of the included studies³⁷ had a longer follow-up period of 84 months. However, the overall pooled result was not substantially altered

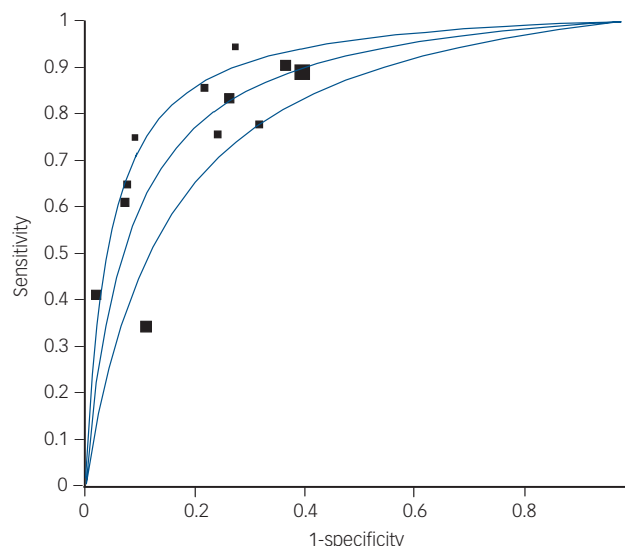


Fig. 1 Summary receiver operator characteristic (ROC) curve for ultra-high-risk criteria.

Symmetric summary ROC: area under curve (AUC) = 0.8537, s.e. = 0.232; $Q^* = 0.7847$, s.e. = 0.0221.

by the inclusion or exclusion of this study. Table 1 shows the diagnostic attributes for ultra-high-risk prodromal criteria against a standardised diagnosis of schizophrenia (DSM-IV), whereas Table 2 shows the diagnostic performance of ultra-high-risk criteria at varying cut-off points.

Properties of basic-symptoms criteria in predicting schizophrenia

We identified only one study (160 participants: 79 confirmed with schizophrenia) reported in two publications.^{24,36} The study had a sensitivity of 0.97 (95% CI 0.91–1.00), specificity of 0.59 (95% CI 0.48–0.70) and an average follow-up period of 9.6 years. The likelihood ratio for a positive test was 2.39 (95% CI 1.84–3.12) and the likelihood ratio for a negative test was 0.04 (95% CI 0.01–1.17). The cut-off point for these results was the presence of at least 1 of 66 basic symptoms at baseline. The study results at two different cut-off points were also available. The second cut-off point was the presence of at least one of the ten cognitive perceptive (COPER) basic symptoms.³⁶ This had a sensitivity of 0.87, a specificity of 0.54, a positive likelihood ratio of 1.9 and a

Study	n	Follow-up Months	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Yung <i>et al</i> 2003 ²⁵	48	12	0.65 (0.41–0.83)	0.92 (0.73–0.99)	8.45 (2.15–33.25)	0.38 (0.21–0.69)
Yung <i>et al</i> 2004 ²⁶	104	12	0.61 (0.44–0.76)	0.93 (0.83–0.93)	8.37 (3.44–20.10)	0.42 (0.28–0.60)
Mason <i>et al</i> 2004 ²⁷	74	>12	0.84 (0.67–0.93)	0.86 (0.70–0.95)	6.2 (2.7–14.18)	0.2 (0.09–0.40)
Yung <i>et al</i> 2005 ²⁸	150	6	0.83 (0.36–0.99)	0.74 (0.65–0.80)	3.16 (2.01–4.95)	0.22 (0.04–1.30)
Yung <i>et al</i> 2008 ²⁹	292	24	0.91 (0.68–0.98)	0.63 (0.57–0.69)	2.45 (1.99–3.02)	0.15 (0.04–0.57)
Woods <i>et al</i> 2009 ³³	638	30	0.89 (0.81–0.94)	0.60 (0.56–0.64)	2.44 (1.98–2.53)	0.18 (0.10–0.32)
Miller <i>et al</i> 2003 ³	34	24.7	1.00 (0.60–1.00)	0.73 (0.45–0.91)	3.75 (1.62–8.62)	0
Lemos <i>et al</i> 2006 ³⁰	30	12	0.75 (0.47–0.92)	0.91 (0.64–0.99)	10.5 (1.5–70.9)	0.27 (0.11–0.64)
Lencz <i>et al</i> 2003 ³⁵	34	24	0.78 (0.40–0.96)	0.68 (0.46–0.84)	2.43 (1.24–4.75)	0.33 (0.09–1.15)
Riecher-Rossler <i>et al</i> 2009 ³⁷	53	84	0.81 (0.58–0.94)	0.78 (0.60–0.91)	3.7 (1.86–7.36)	0.16 (0.06–0.40)
Ruhrmann <i>et al</i> 2010 ³⁴	183	18	0.42 (0.25–0.59)	0.98 (0.94–1.00)	19.9 (6.06–65.4)	0.60 (0.45–0.80)
Cannon <i>et al</i> 2008 ³¹	291	30	0.34 (0.25–0.44)	0.89 (0.84–0.93)	3.1 (1.90–5.01)	0.74 (0.64–0.86)

Table 2 Diagnostic performance of ultra-high-risk criteria at varying cut-off points^a

Study and cut-off/result description	Sensitivity	Specificity	Predictors/index tests
Yung <i>et al</i> 2003, ²⁵ presence of at least one of four potential predictors	0.65	0.93	SCID for DSM-IV, ⁴⁰ QLS, ⁴¹ BPRS, ⁴² SANS, ⁴³ HRSD, ⁴⁴ HRSA, ⁴⁵ MRS ⁴⁶
Yung <i>et al</i> 2004 ²⁶ Trait + attenuated symptoms	0.60	0.93	SCID for DSM-IV, QLS, BPRS, SANS, HRSD, HRSA, MRS, GAF ⁴⁷
Presence of at least one of four potential predictors	0.61	0.93	
Mason <i>et al</i> 2004 ²⁷ Unusual thought content, magical ideation	0.84	0.84	PSA, ⁴⁸ QLS, SRE, ⁴⁹ BPRS, SANS, GAF, HRSD, HRSA
Schizotypal personality disorder	0.76	0.76	
Yung <i>et al</i> 2005, ²⁸ presence of at least one of four potential predictors	0.83	0.74	SCID for DSM-IV, CAARMS ²
Yung <i>et al</i> 2006, ³² 2008, ²⁹ presence of at least one of four potential predictors	0.92	0.62	SCID for DSM-IV, CAARMS, GAF
Woods <i>et al</i> 2009, ³³ ultra-high-risk positive	0.89	0.60	SIPS, ³ SOPS, ³ GAF
Miller <i>et al</i> 2003, ³ SIPS positive	1.0	0.73	SIPS, SCID for DSM-IV, SOPS, GAF, COPS, POPS
Lemos <i>et al</i> 2006 ³⁰ Positive symptoms	0.75	0.91	SCID for DSM-IV, SOPS, SIPS
Negative symptoms	1.00	0.96	
Disorganised symptoms	1.00	0.86	
General symptoms	0.75	0.93	
Lencz <i>et al</i> 2003 ³⁵ Total score of 10+	0.70	0.72	SCID for DSM-IV, SOPS, SIPS
Total score 15+	0.79	0.69	
Highest item score 5	0.80	0.59	
Highest item score >3	0.81	0.78	
Riecher-Rössler <i>et al</i> 2009 ³⁷ Positive symptoms (BPRS) suspiciousness	0.70	0.72	BSIP, ⁵⁰ BPRS, SANS
Negative symptoms (SANS) alogia, anhedonia-asociality	0.79	0.69	
Neuropsychology	0.80	0.59	
Combined model	0.81	0.78	
Ruhrmann <i>et al</i> 2010, ³⁴ SIPS positive score >16	0.42	0.98	SIPS, GAF
Cannon <i>et al</i> 2008, ³¹ genetic risk and GAF reduction of ≥10%, unusual thought content, and paranoid ideation	0.34	0.89	SIPS, GAF

SCID for DSM-IV, Structured Clinical Interview for DSM-IV; QLS, Quality of Life Scale; BPRS, Brief Psychiatric Rating Scale; SANS, Scale for the Assessment of Negative Symptoms; HRSD, Hamilton Rating Scale for Depression; HRSA, Hamilton Rating Scale for Anxiety; MRS, Mania Rating Scale; PSA, Premorbid Social Adjustment scale; SRE, Schedule of Recent Experience; GAF, Global Assessment of Functioning; CAARMS, Comprehensive Assessment of at Risk Mental States; SIPS, Structured Interview for Prodromal Syndromes; SOPS, Scale of Prodromal Symptoms; COPS, Criteria of Prodromal Syndromes; POPS, Presence of Psychotic Syndromes; BSIP, Basel Screening Instrument for Psychosis.
a. Ruhrmann *et al* in addition to using ultra-high-risk criteria also assessed cognitive disturbance (COGDIS).

negative likelihood ratio of 0.2.³⁶ The third cut-off point was the presence of at least one of nine cognitive disturbance (COGDIS) basis symptoms.³⁶ The sensitivity, specificity, positive likelihood ratio and negative likelihood ratio of this criteria was 0.67, 0.83, 3.9 and 0.4 respectively.

Discussion

Key findings

We found 12 predictive diagnostic studies with 1918 participants where a diagnostic gold standard was independently applied. Ultra-high-risk criteria were able to correctly predict schizophrenia (sensitivity 81%) while being able to exclude this condition with some certainty (specificity 67%). The AUC for the summary ROC was 0.85, suggesting that the ultra-high-risk criteria can predict the conversion to schizophrenia with moderate precision.³⁹

We only identified one study²⁴ investigating the diagnostic properties of basic-symptoms criteria against a DSM-IV gold standard. The study had a very good sensitivity of 0.97, moderate specificity of 0.59 and a follow-up period of over 9.6 years.

Strength and limitations

An interesting finding in relation to ultra-high-risk criteria is the observation period. The identified studies had an average

follow-up period of 18.6 months; a recent retrospective study³⁸ estimated that the pre-psychotic period can continue for up to 7 years. This suggests that some people identified as false positives by the prodromal criteria may remain at risk of developing schizophrenia in the future. Treating those who remain at high risk of developing schizophrenia at the end of the study period as false positives may have the effect of underestimating the diagnostic properties of prodromal criteria. The argument that false positives may represent individuals who remain at risk of developing schizophrenia is only true for a true prodrome. True schizophrenia prodrome is defined as a state inevitably and continuously progressing into the manifest of schizophrenia disorder.¹³ However, this argument is not true for self-restricting, non-progressive episodes of an 'at risk mental state' (also known as outpost syndrome) that have recently been reported to frequently occur in adolescents.^{13,51} Outpost syndrome therefore represents true false positives (people identified by the test as at high risk but who do not convert to schizophrenia). The prevalence of these self-restricting episodes without subsequent development of psychosis is not known.¹³

The strength of the identified study using basic-symptoms criteria was that it had a long follow-up period of 9.6 years and an adequate sample size (160 participants). The follow-up period of 9.6 years allows for a sufficient time to observe the participants at 'at high risk' throughout the pre-psychotic phase, unlike those

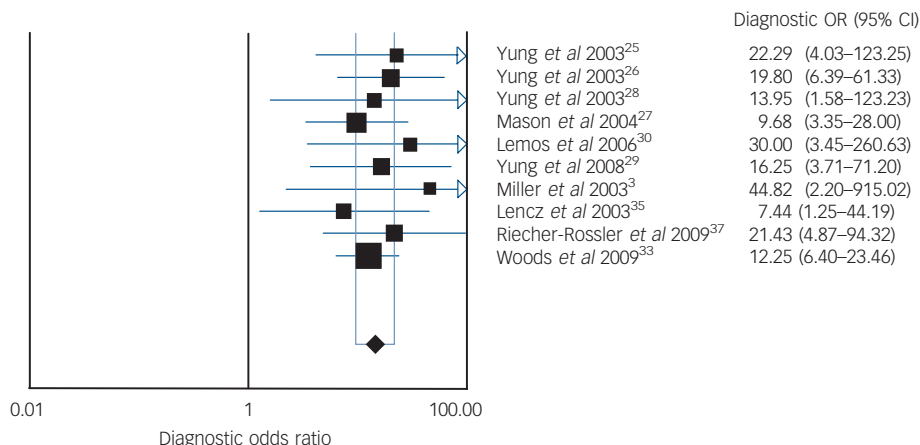


Fig. 2 Diagnostic odds ratio (OR) for ultra-high-risk criteria.

Random effects model: pooled diagnostic OR = 14.32 (95% CI 9.57–21.44). Cochran $Q = 3.15$, d.f. = 9, $P = 0.9581$. Inconsistency $I^2 = 0.0\%$, $\tau^2 = 0.0000$.

using ultra-high-risk criteria, which had an average follow-up of 18.6 months. The psychometric attributes of basic-symptoms criteria need to be subject to a much wider range of validation studies across different clinical settings and patient populations before its validity can be assumed.

The methodological quality of several studies was poor according to our main chosen criterion of masked application of a diagnostic gold standard. All studies^{3,24–31,33–35,37} did not specify whether investigators were masked to the results of the other test (index test or reference test; item 11 of the STARD checklist). All identified prospective studies^{3,24–31,33–35,37} were from specialised early detection assessment settings, which suggests selection bias and therefore makes it difficult to generalise these studies to the general population.

Implications

Findings from this systematic review suggest that ultra-high-risk and basic-symptoms criteria are valid tools in predicting the future development of schizophrenia among the 'at-risk population'. More long-term studies need to be undertaken using the ultra-high-risk criteria to eliminate the possibility of false positives given the possible long duration of a prodromal state. The predictive validity of basic-symptoms criteria was supported by one study; we recommend future research applying the basic-symptoms criteria in different settings and cultures.

Jeffer Chuma, MSc, BSc, RMN, **Prem Mahadun**, MBBS, MSc Psychiatry, MRCPsych, Trafford Crisis Resolution and Home Treatment Team, Greater Manchester West NHS Foundation Trust, Manchester, UK

Correspondence: Jeffer Chuma, Greater Manchester West NHS Foundation Trust, Trafford Crisis Resolution and Home Treatment Team, 71A Chapel Road, Sale, Manchester M33 7EG, UK. Email: jefta.chuma@trafford.nhs.uk

First received 9 Sep 2010, final revision 20 Feb 2011, accepted 27 Apr 2011

Acknowledgements

We are grateful to Professor Simon Gilbody for comments on an earlier draft of this manuscript. We also thank Professor Martin Bland for his comments and assistance with statistical methods used in the analysis of this study.

References

- Knowles L, Sharma T. Identifying vulnerability markers in prodromal patients: a step in the right direction. *CNS Spectrums* 2004; **9**: 594–602.
- Yung AR, Phillips LJ, Yuen HP, McGorry PD, Ward J, Donovan K, et al. *Comprehensive Assessment of at Risk Mental States (CAARMS)*, January 2002. PACE Clinic, Department of Psychiatry, University of Melbourne, 2002.
- Miller TJ, McGlashan TH, Rosen JL, Cadenhead K, Ventura J, McFarlane W, et al. Prodromal assessment with structured interview for prodromal syndromes and scale of prodromal symptoms: predictive validity, interrater reliability and training to reliability. *Schizophr Bull* 2003; **29**: 703–15.
- Phillips LJ, Yung AR, McGorry PD. Identification of young people at risk of psychosis: validation of personal assessment and crisis evaluation intake criteria. *Aust N Z J Psychiatry* 2000; **34**: 164–9.
- Olsen KA, Rosenbaum B. Prospective investigations of prodromal state of schizophrenia: assessment instruments. *Acta Psychiatr Scand* 2006; **113**: 273–82.
- Gross G, Huber G, Klosterkotter J. *Bonn Scale for the Assessment of Basic Symptoms – BSABS*. Springer, 1987.
- McGlashan TH, Miller TJ, Woods SW, Rosen JL, Hoffman RE, Davidson L. *Structured Interview for Prodromal Syndromes (Version 4.0)*. Yale School of Medicine, 2003.
- Marshall M, Lewis S, Lockwood A. Association between duration of untreated psychosis and outcome in cohorts of first episode patients: a systematic review. *Arch Gen Psychiatry* 2005; **62**: 975–83.
- Perkins D, Gu H, Botveta K. Relationship between duration of untreated psychosis and outcome in first episode schizophrenia: a critical review and meta-analysis. *Am J Psychiatry* 2005; **162**: 1785–804.
- McGlashan TH, Addington J, Cannon T, Heinimaa M, McGorry P, O'Brien M, et al. Recruitment and treatment practices for help seeking prodromal patients. *Schizophr Bull* 2007; **33**: 1–2.
- Haroun N, Dunn L, Haroun A, Cadenhead KS. Risk and protection in prodromal schizophrenia: ethical implications for clinical practice and future research. *Schizophr Bull* 2006; **32**: 166–78.
- Warner R. Problems with early and very early intervention in psychosis. *Br J Psychiatry* 2005; **187**: s104–7.
- Ruhrmann S, Schultze-Luther F, Klosterkotter J. Probably at-risk, but certainly ill – advocating the introduction of a psychosis spectrum disorder in DSM-V. *Schizophr Res* 2010; **120**: 23–37.
- Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak* 2007; **7**: 16.
- Deville WL, Bunnix F, Bouter LM, Montori VM, De Vet HC, Van Der DA, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2003; **2**: 9.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (4th edn) (DSM-IV)*. APA, 1994.

- 17 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD Statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; **138**: w11–12.
- 18 Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; **3**: 25.
- 19 NHS Centre for Reviews and Dissemination. *Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for those Carrying out or commissioning Reviews*. NHS Centre for Reviews and Dissemination, University of York, 2001.
- 20 Sackett DL, Haynes RB. Evidence base of clinical diagnosis: the architecture of diagnostic research. *BMJ* 2002; **324**: 539–41.
- 21 Zamora J, Abraira V, Muriel A, Khan KS, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006; **6**: 31.
- 22 Walter SD. Properties of the summary receiver-operating characteristic (sROC) curve for diagnostic test data. *Stat Med* 2002; **21**: 1237–56.
- 23 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**: 557–60.
- 24 Klosterkötter J, Hellmich M, Steinmeyer EM, Schultze-Lutter F. Diagnosing schizophrenia in the initial prodromal phase. *Arch Gen Psychiatry* 2001; **58**: 158–64.
- 25 Yung AR, Phillips LJ, Yuen HP, Francey SM, McFarlane CA, Hallgren M, et al. Psychosis prediction: 12 months follow up of at risk (prodromal) group. *Schizophr Res* 2003; **60**: 21–32.
- 26 Yung AR, Phillips LJ, Yuen HP, McGorry PD. Risks factors for psychosis in an ultra-high risk group: psychological and clinical features. *Schizophr Res* 2004; **67**: 131–42.
- 27 Mason O, Startup M, Halpin S, Schall U, Conrad A, Carr V. Risk factors for transition for first episode psychosis among individuals with 'at-risk mental states'. *Schizophr Res* 2004; **71**: 2–3.
- 28 Yung AR, Phillips LJ, Yuen HP, McGorry PD, Kelly D, Dell'olio M, et al. Mapping the onset of psychosis: the comprehensive assessment of at risk mental states. *Aust N Z J Psychiatry* 2005; **39**: 964–71.
- 29 Yung AR, Nelson B, Stanford C, Simmons MB, Cosgrave EM, Killackey E, et al. Validation of 'prodromal' criteria to detect individuals at ultra high risks of psychosis: 2 year follow up. *Schizophr Res* 2008; **105**: 10–7.
- 30 Lemos S, Vanilla O, Fernandez P, Ortega JA, Garcia P, Gutierrez A, et al. Validez predictiva de la escala de sintomas prodromicos (SOPS). [Predictive validity of the Scale of Prodromal Symptoms (SOPS).] *Actas Esp Psiquiatr* 2006; **34**: 216–23.
- 31 Cannon TD, Cadenhead K, Cornblatt B, Woods SW, Addington J, Walker E, et al. Prediction of psychosis in youth at high clinical risk: a multisite longitudinal study in North America. *Arch Gen Psychiatry* 2008; **65**: 28–37.
- 32 Yung AR, Stanford C, Cosgrave E, Killackey E, Phillips L, Nelson B, et al. Testing the ultra high risk (prodromal) criteria for the prediction of psychosis in a clinical sample of young people. *Schizophr Res* 2006; **84**: 57–66.
- 33 Woods SW, Addington J, Cadenhead KS, Cannon TD, Cornblatt BA, Heinssen R, et al. Validity of the prodromal risk syndrome for first psychosis: findings from the North American Prodrome Longitudinal Study. *Schizophr Bull* 2009; **35**: 894–908.
- 34 Ruhrmann S, Schultze-Lutter F, Salokangas RKR, Heinimaa M, Linszen D, Dingeramn P, et al. Prediction of psychosis in adolescents and young adults at high risk: results from the prospective European prediction of psychosis study. *Arch Gen Psychiatry* 2010; **67**: 241–51.
- 35 Lencz T, Smith CW, Auther AM, Correl CU, Cornblatt BA. The assessment of 'prodromal Schizophrenia: unresolved issues and future directions. *Schizophr Bull* 2003; **29**: 717–28.
- 36 Schultze-Lutter F, Ruhrmann S, Klosterkötter J. Can schizophrenia be predicted phenomenologically? In *Evolving Psychosis: Different Stages, Different Treatments* (eds JO Johannesen, BV Martidale, J Cullberg): 104–23. Routledge, 2006.
- 37 Riecher-Rössler A, Pflueger MO, Aston J, Borgwardt SJ, Brewer WJ, Gschwandtner U, et al. Efficacy of using cognitive status in predicting psychosis: a 7-year follow up. *Biol Psychiatry* 2009; **66**: 1023–30.
- 38 Schultze-Lutter F, Ruhrmann S, Berning J, Maier W, Klosterkötter J. Basic symptoms and ultrahigh risk criteria: symptom development in the initial prodromal state. *Schizophr Bull* 2010; **36**: 182–91.
- 39 Streiner D, Norman G. *Health Measurement Scales: A Practical Guide to their Development and Use (3rd edn)*. Oxford University Press, 2003.
- 40 First MB, Spitzer RL, Gibbon M, Williams JB. *Structured Clinical Interview for DSM-IV Axis I Disorders*. New York State Psychiatric Institute, 1995.
- 41 Heinrichs D, Hanlon T, Carpenter W. The quality of life scales an instrument for rating the schizophrenia deficit syndrome. *Schizophr Bull* 1984; **10**: 388–98.
- 42 McGorry PD, Goodwin RJ, Stuart GW. The development use and reliability of the brief psychiatric rating scale (nursing modification) an assessment procedure for the nursing team in clinical setting. *Compr Psychiatry* 1988; **29**: 575–87.
- 43 Andreasen N. Negative symptoms of schizophrenia: definition and reality. *Arch Gen Psychiatry* 1982; **39**: 784–8.
- 44 Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; **23**: 56–61.
- 45 Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol* 1959; **32**: 50–5.
- 46 Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability validity and sensitivity. *Br J Psychiatry* 1978; **133**: 429–35.
- 47 Hall R. Global assessment of functioning: a modified scale. *Psychosom* 1995; **36**: 267–75.
- 48 Cannon-Spoor HE, Potkin SO, Wyatt RJ. Measurement of premorbid adjustment in chronic schizophrenia. *Schizophr Bull* 1982; **8**: 471–84.
- 49 Amundson ME, Hart CA, Holmes TH. *Manual for the Schedule of Recent Experience (SRE)*. University of Washington, 1981.
- 50 Riecher-Rössler A, Aston J, Ventura J, Merlo M, Borgwardt S, Gschwandtner U. The Basel Screening Instrument for Psychosis (BSIP): development, structure, reliability and validity. *Fortschr Neurol Psychiatr* 2008; **76**: 207–16.
- 51 Simon AE, Umbricht D. High remission rates from an initial ultra-high risk state for psychosis. *Schizophr Res* 2010; **116**: 168–72.

