Review

Meta-analytical prognostic accuracy of the Comprehensive Assessment of at Risk Mental States (CAARMS): The need for refined prediction

D. Oliver [a,*], M. Kotlicka-Antczak [b], A. Minichino [c], G. Spada [a], P. McGuire [d,e,f], P. Fusar-Poli [a,e,f]

[a] Early Psychosis: Interventions & Clinical-detection (EPIC) lab, Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, United Kingdom
[b] Medical University of Lodz, Department of Affective and Psychotic Disorders, Lodz, Poland
[c] Department of Neurology and Psychiatry, Sapienza University of Rome, Rome, Italy
[d] Department of Psychosis Studies, IoPPN, King's College London, London SE5 8AF, United Kingdom
[e] OASIS Service, South London and the Maudsley NHS National Health Service Foundation Trust, London, United Kingdom
[f] National Institute for Health Research (NIHR) Biomedical Research Centre for Mental Health, IoPPN, King's College London, SE5 8AF, United Kingdom

ARTICLE INFO

ABSTRACT

Primary indicated prevention is reliant on accurate tools to predict the onset of psychosis. The gold standard assessment for detecting individuals at clinical high risk (CHR-P) for psychosis in the UK and many other countries is the Comprehensive Assessment for At Risk Mental States (CAARMS). While the prognostic accuracy of CHR-P instruments has been assessed in general, this is the first study to specifically analyse that of the CAARMS. As such, the CAARMS was used as the index test, with the reference index being psychosis onset within 2 years. Six independent studies were analysed using MIDAS (STATA 14), with a total of 1876 help-seeking subjects referred to high risk services (CHR-P+: $n = 892$; CHR-P–: $n = 984$). Area under the curve (AUC), summary receiver operating characteristic curves (SROC), quality assessment, likelihood ratios, and probability modified plots were computed, along with sensitivity analyses and meta-regressions. The current meta-analysis confirmed that the 2-year prognostic accuracy of the CAARMS is only acceptable (AUC = 0.79 95% CI: 0.75–0.83) and not outstanding as previously reported. In particular, specificity was poor. Sensitivity of the CAARMS is inferior compared to the SIPS, while specificity is comparably low. However, due to the difficulties in performing these types of studies, power in this meta-analysis was low. These results indicate that refining and improving the prognostic accuracy of the CAARMS should be the mainstream area of research for the next era. Avenues of prediction improvement are critically discussed and presented to better benefit patients and improve outcomes of first episode psychosis.

1. Introduction

Psychosis is a severe psychiatric condition and there is limited evidence that treatments are successful in improving patients' functioning once the disorder is established [1]. Intervening in the earlier phases is therefore the only viable possibility to substantially alter the course of the disorder [2,3]. Within early intervention, a key focus for improving the outcome has been primary indicated prevention [2,4,5]. Primary indicated prevention allows for early intervention for those at clinical high risk of developing psychosis (CHR-P), with greater scope for improving outcomes. To do this effectively, the first necessary step is to reach an accurate, robust prognostic identification of individuals meeting CHR-P criteria who will subsequently develop psychosis or not. Ideally, all subjects who will actually develop psychosis should be classified as "at risk" (CHR-P+) while those not developing an established psychosis should be classified as "not at risk" (CHR-P–). These key concepts involved in prognostic reasoning in the CHR-P have been detailed and presented in a recent paper by our group [6].

Prognostic prediction is used in many branches of medicine to identify individuals who may develop a particular disease [7]. For example, fasting glucose, oral glucose tolerance test and glycated

* Corresponding author.
E-mail address: dominic.a.oliver@kcl.ac.uk (D. Oliver).

haemoglobin are used to detect individuals at high risk for developing diabetes (pre-diabetes or intermediate hyperglycaemia) [8] and systolic blood pressure and ratio of total serum cholesterol to high density lipoprotein cholesterol levels are used to detect individuals at high risk for developing cardiovascular disease [9]. However, unlike these other fields, there are no biological tests to assess the risk of developing mental disorders [10], which is instead reliant on semi-structured CHR-P psychometric interviews, such as the CAARMS (Comprehensive Assessment for At Risk Mental States) [11]. Recently, the CAARMS has become the mainstream tool to detect CHR-P individuals in the UK, recommended by international bodies, such as NICE [12]. Therefore, understanding its exact psychometric properties is of paramount clinical relevance. The CAARMS shows excellent inter-rater reliability when performed by trained raters (0.85) [13]. However, its prognostic accuracy is uncertain. A recent meta-analysis by our lab [14] investigated the prognostic accuracy of CHR-P instruments, showing generally excellent prognostic performance of these instruments. However, CHR-P tools were grouped together including the CAARMS [11], the SIPS (Structured Interview for Prodromal Syndromes) [15] and the SPI-A (Schizophrenia Proneness Instrument-Adult Version) [16]. This was due to the fact that there were not enough studies contributing data to assess the meta-analytical prognostic accuracy of the CAARMS specifically. Given the marked differences between the CAARMS and other CHR-P instruments [17], in particular with respect to the functional deterioration criterion [18], it is possible that the previously reported meta-analytical prognostic accuracy is not completely accurate. In addition, the previous meta-analysis combined multiple follow-up time points, and even though meta-regressions of this variable found no significant effect, validity of the prognostic accuracy results would be improved by using a more defined and consistent follow-up time [14].

The current study tackles these caveats and advances knowledge in the psychometric properties of the CAARMS. We capitalize on recently published CAARMS studies reporting useful and innovative meta-analytical data to conduct a meta-analytical prognostic accuracy analysis of the CAARMS at two-year follow-up. This is the period of time during which most transitions to psychosis occur [19]. The results will hopefully support the refinement of psychosis prediction and therefore facilitate indicated primary prevention in CHR-P individuals.

2. Methods

2.1. Search strategy

Two investigators (DO, PFP) conducted a two-step literature search. At a first step, the Web of Knowledge database was searched, incorporating both the Web of Science and Medline. The search was extended until August 2017, only including abstracts in English. The electronic research adopted several combinations of the following keywords: "at risk mental state", "psychosis risk", "prodrome", "prodromal psychosis", "ultra-high risk", "high risk", "help-seeking", "diagnostic accuracy", "sensitivity", "specificity", "psychosis prediction", "psychosis onset". The second step involved the use of Scopus to investigate citations of previous systematic reviews on transition outcomes in CHR-P subjects and a manual search of the reference lists of the retrieved articles.

Articles identified through these two steps were then screened for the selection criteria on the basis of abstract reading. The articles surviving this selection were assessed for eligibility on the basis of full text reading. To achieve a high standard of reporting, we adopted the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) checklist [20].

2.2. Selection criteria

Studies were eligible for inclusion if:

- they were reported in original articles, written in English;
- they had used the CAARMS (index test) in the same pool of referrals;
- they had followed up both CHR-P+ and CHR-P– subjects for psychosis onset (reference index) using established international diagnostic manuals (ICD or DSM);
- they had reported sufficient prognostic accuracy data at 2-year follow-up.

With respect to this last point, when data were not directly presented, they were indirectly extracted from associated data. Additionally, we contacted all corresponding authors to request additional data when needed.

We excluded:

- abstracts, reviews, articles in a language other than English;
- studies in which interviews were not conducted in the same pool of referrals or that used an external CHR-P group of healthy controls;
- studies with overlapping datasets.

In case of multiple publications deriving from the same study population, we selected the article reporting the largest and most recent data set. The literature search was summarized according to PRISMA guidelines [21].

2.3. Recorded variables

Data extraction was independently performed by two investigators (DO, PFP). Data included author, year of publication, characteristics of subject samples (baseline sample sizes, mean age and age range, proportion of females), diagnostic criteria used at follow-ups to assess the psychotic outcome, prognostic accuracy data (number of true and false positives, true and false negatives or associated data) and quality assessment conducted with the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) checklist [22].

2.4. Statistical analysis

The statistical analysis followed the Cochrane Guidelines for Systematic Reviews of Diagnostic Test Accuracy, Version 1.0 [23] and the Methods Guide for Authors of Systematic Reviews of Medical Tests by the Agency for Healthcare Research and Quality (chapter 8) [24]. Evaluating test accuracy requires knowledge of two quantities: the test's sensitivity (Se) and specificity (Sp). Meta-analysis methods for diagnostic test accuracy thus have to deal with two summary statistics simultaneously rather than one [23]. Methods for undertaking analyses, which account for both Se and Sp, the relationship between them, and the heterogeneity in test accuracy, require fitting advanced hierarchical random effects models [23].

For each study, we constructed a two-by-two table, which included true positive, false positive, true negative, and false negative values. The baseline sample size was conservatively used as the base reference.

Data were then analysed with MIDAS (Meta-analytical Integration of Diagnostic Accuracy Studies) [25], a comprehensive program of statistical and graphical routines for undertaking meta-analysis of diagnostic/prognostic test performance in STATA 14 software [26]. The index tests of CHR-P status (CHR-P+ or CHR-P–) and reference tests of transition to psychosis according to international

diagnostic manuals (ICD or DSM as gold standard) were dichotomous.

Primary data synthesis was performed within the bivariate mixed-effects regression framework for the logit transforms of Se and Sp. In addition to accounting for study size, the bivariate model estimates and incorporates the intrinsic negative correlation that may arise between Se and Sp within studies (threshold effect) [27], as a result of differences in the test threshold between studies [28]. The bivariate model allows for heterogeneity beyond chance as a result of clinical and methodological differences between studies [28].

We estimated the summary Se and Sp and the hierarchical SROC (summary receiver operator characteristic) curves [23,32]. A SROC graph across each predictor, with the y-axis representing the predictor's Se and the x-axis representing 1–specificity, was used to plot a 95% confidence region and a 95% prediction region around the summary estimates to illustrate the precision with which the summary values were estimated (confidence ellipse of a mean), and to show the amount of between-study variation (prediction ellipse; the likely range of values for a new study). We also estimated the AUC (area under the curve). The AUC serves as a global measure of test performance. Values in the range of 0.9–1 are considered outstanding, between 0.8 and 0.9 are considered excellent, between 0.7 and 0.8 are considered acceptable [29].

Heterogeneity across studies was assessed using the I^2, with values of 25%, 50% and 75% representing mild, moderate and severe inconsistency, respectively [30]. Within MIDAS, forest plots and heterogeneity statistics can be created for each test performance parameter individually or may be displayed as paired plots. Meta-regressions were used to examine the influence of mean age, gender (% females), sample size, and quality assessment (QUADAS) on meta-analytical estimates. Furthermore, we investigated the prognostic accuracy difference between CAARMS based studies and studies employing the SIPS, as detected in the previous meta-analysis [14]. To control for biases associated with imbalanced datasets [31], we further tested the impact of the proportion of CHR-P+ subjects in the overall samples. The meta-regressions were used if there was substantial heterogeneity ($I^2 > 50\%$) [32] and when more than 10 studies were available.

Sensitivity analyses (i.e., exclusion of outliers and rerunning of the model) were conducted to further explore heterogeneity. We did not test publication bias [33], because no proven statistical method exists for this type of meta-analysis [34].

In a second step, we employed the probability-modifying plot to estimate the clinical or patient-relevant utility of the CAARMS in subjects seeking help at CHR-P services.

The clinical utility was evaluated using the positive and negative likelihood ratios (LR+ and LR–) to calculate post-test probability (post-TP) based on Bayes' theorem (with pre-test probability, pre-PT, being the prevalence of the condition in the target population), as follows: post-TP = LR × pre-TP/[(1 – pre-TP) + (pre-TP × LR)] [27]. Specifically, the probability-modifying plot [25] is a graphical sensitivity analysis of the test's predictive values across a baseline psychosis risk continuum in people seeking help at CHR-P services. It depicts separate curves for positive and negative tests and uses general summary statistics (i.e., unconditional positive and negative predictive values, NPV and PPV, which permit underlying psychosis risk heterogeneity) to evaluate the effect of the CHR-P assessment on predictive values [35]. The pre-TP probability of psychosis risk in subjects seeking help at early detection services was computed in the current dataset as the proportion of subjects developing psychosis on the total baseline sample (CHR-P+ plus CHR-P–) [25].

Statistical tests were two-sided and statistical significance was defined as P values < 0.05.

3. Results

3.1. Database

The literature search produced 6 independent studies [36–41] that met inclusion criteria with a total of 1876 subjects (CHR+: n = 892; CHR–: n = 984) referred to clinical high risk services. The dataset was balanced with CHR+ individuals composing 47.5% of the total subjects. The characteristics of the studies are reported in the Table 1 while the PRISMA diagram is depicted in Fig. 1. The MOOSE checklist is reported in the eTable 1. The detailed QUADAS assessment is reported in the eTable 2 and eFig. 1.

3.2. Prognostic accuracy of the CAARMS at 2-years

The summary meta-analytical estimate of Se at 2 years (0.86, 95% CI = 0.76–0.92) was outstanding with the 2-year AUC (0.79, 95% CI = 0.75–0.83) acceptable but the estimate of 2-year Sp (0.55, 95% CI = 0.48–0.63) was poor (Fig. 2). There was severe heterogeneity present in this analysis (I^2 = 93.28%, 95% CI = 89.42–97.15%).

3.3. Clinical utility of the CAARMS at 2 years

The 2-year psychosis transition risk in the 1876 subjects was 0.09 (95% CI = 0.05–0.13). On the basis of the prior distribution, the continuous relationship between pre-TP and post-TP probability is summarized in Fig. 3. Being CHR-P+ was associated with a 0.16 (95% CI = 0.10–0.22) risk of developing psychosis within 2 years, yet a small LR+ of just 1.9 (95% CI = 1.5–2.4) while being CHR-P– was associated with a 0.03 (95% CI = 0.02–0.05) risk of transition to psychosis with a moderate LR– of 0.25 (95% CI = 0.13–0.48).

3.4. Meta-regressions and sensitivity analyses

Sensitivity analysis suggested that one study [41] was influential with a Cook's distance > 1. While we hypothesised this was due to the study reporting 0 false negatives, we were unable to test the effect of false negatives through meta-regression due to low number of studies. Similarly, we were unable to perform meta-regressions for age, gender, QUADAS score or sample size as there were fewer than 10 studies contributing data. As indicated in the methods, we were able to perform a meta-regression comparing the prognostic accuracy of the CAARMS vs. that of the SIPS, using the studies reporting SIPS data [42–46] only as identified in our previous study (n = 5, CHR+: n = 783; CHR–: n = 360). As indicated in Fig. 4, Se was significantly higher (P < 0.001) for the SIPS (n = 5, mean = 0.95, 95% CI 0.91–0.99) compared to the CAARMS (n = 6, mean = 0.87, 95% CI 0.79–0.96), while Sp was comparably (P = 0.27) low in the SIPS (n = 5, mean = 0.45, 95% CI = 0.38–0.53) and in the CAARMS (n = 6, mean = 0.55, 95% CI 0.48–0.62).

4. Discussion

This is the first meta-analysis specifically investigating the prognostic accuracy of the CAARMS for the prediction of psychosis. We found 6 studies that investigated prognostic accuracy of the CAARMS at two-year follow-up, which contributed a relatively large database of 1876 subjects overall, with 892 considered CHR-P+ and 984 CHR-P–. Prognostic accuracy of the CAARMS in terms of AUC was found to be only acceptable (0.79), mostly mediated by its substantial ability to rule out psychosis (i.e. LR– was relatively

Table 1
Independent studies included in the meta-analysis (studies $n = 6$; 1876 subjects; CHR+: $n = 892$; CHR−: $n = 984$).

Study	QUADAS score (14 = max); exposure to antipsychotics at baseline	Psychosis diagnosis (reference standard)	Age (mean ± SD, range)	Gender (% females)	CHR-P+ (baseline)	CHR-P– (baseline)
1. Yung et al., 2008 [39]	12; yes (N/A)	CAARMS	18.1 (15–24)	51.0	119	173
2. Lee et al., 2013 [41][a]	13; no	DSM-IV	21.6 ± 3.5 (14–29)	39.9	173	494
3. Fusar-Poli et al., 2017 [36]	11.5	ICD-10	23 ± 5.4 (12–44)	43.8	411	299
4. Francesconi et al., 2017 [37]	12	CAARMS	24.3 ± 3.5	47.0	54	62
5. Kotlicka-Antczak et al., 2015 [40][a]	11.5; yes (10.2%)	ICD-10	19.05 ± 3.6 (15–29)	51.1 (c)	94	33
6. Spada et al., 2015 [38]	11; no	DSM-IV	15.8 ± 1.7 (12–17)	47.5	22	18

CAARMS: Comprehensive Assessment of At Risk Mental States; DSM-IV: Diagnostic and Statistical Manual of Mental Disorders; ICD-10: International Classification of Diseases Tenth Revision; N/A: Not available.

[a] Updated follow-up data provided by the authors.
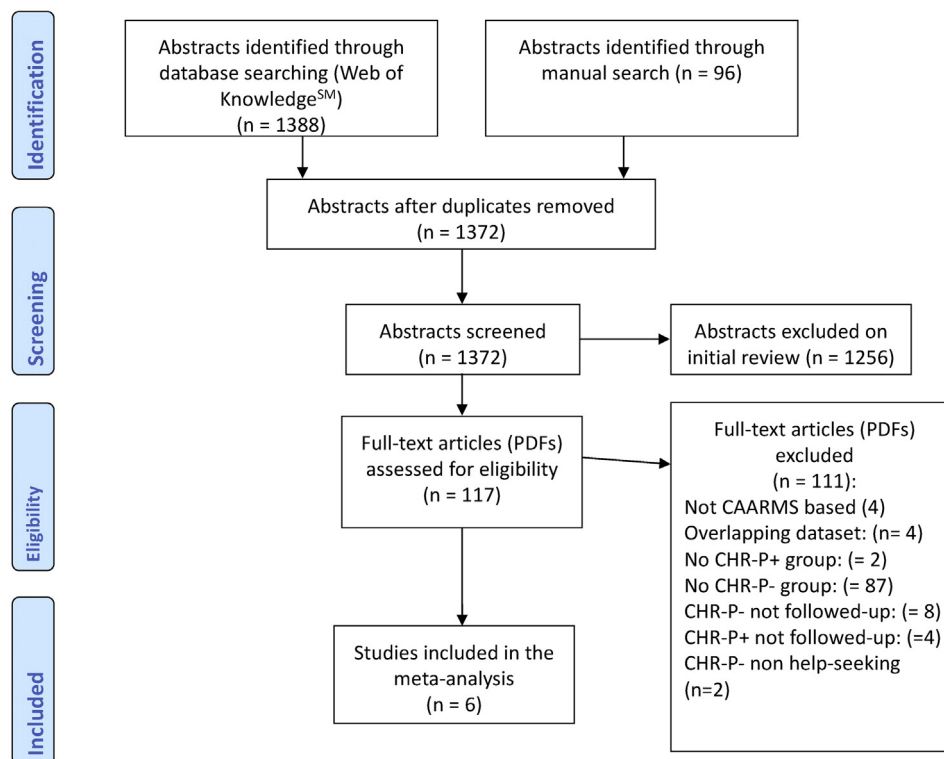
Fig. 1. PRISMA flow chart.

small and Se high). However, this was at the expense of ruling in psychosis (i.e. LR+ was small and Sp was poor). While prognostic accuracy was overall acceptable, this study indicates that refining the prediction of outcomes should be the key priority of future research in this field.

The primary aim of the study was to synthesize available data for the prognostic accuracy of the CAARMS in determining psychosis risk 2 years after young help-seeking subjects presented to CHR-P services. As noted in the introduction, our recent meta-analysis [14] looked into the prognostic accuracy of CHR-P instruments as a collective. The current study advances knowledge indicating that the exact prognostic accuracy of the CAARMS alone is weaker (0.79) than the overall value previously observed when the CHR-P instruments were pooled together (0.90). Although not as outstanding as before, the AUC value here reported is still

considered to be acceptable for a diagnostic test and is comparable to other prognostic tools used in different areas of medicine, such as the AUC = 0.76 attributed to the Cambridge risk score for pre-diabetes [47]. In a similar fashion, we found that the Se (0.86) of the CAARMS alone was less impressive than the Se (0.96) of CHR-P instruments assessed in the previous meta-analysis. Interestingly, there was an apparent minor increase in Sp (0.55 for CAARMS alone compared to 0.47 for CHR-P instruments generally) [14]. The lower AUC compared to the previous general estimate may reflect profound operationalization differences between the CAARMS and the other CHR-P instruments. For example, a comparative analysis between the CAARMS and the SIPS confirmed caseness discrepancies between the two instruments [17], mostly due to different definition of brief limited intermittent psychotic cases, ascertainment of comorbidities [17] and of functional level at intake [18]. To
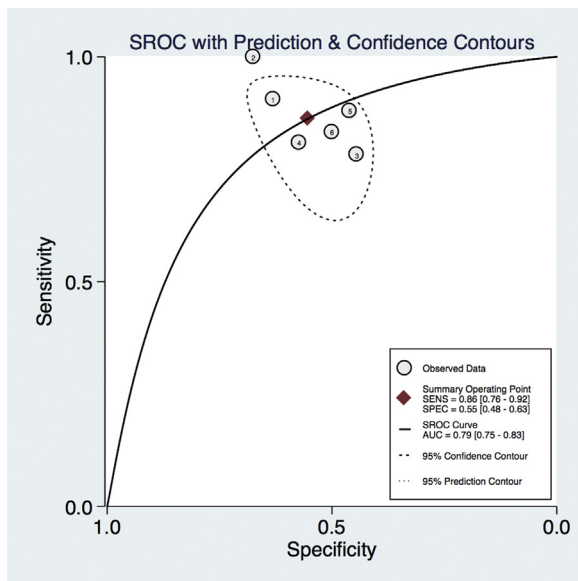
Fig. 2. Meta-analytical summary receiver operating characteristic (SROC) curve of CAARMS assessment. SENS – sensitivity, SPEC – specificity, AUC – area under the curve, 1 – Yung et al., 2008 [39], 2 – Lee et al., 2013 [12], 3 – Fusar-Poli et al., 2017 [2], 4 – Francesconi et al., 2017 [37], 5 – Kotlicka-Antczak et al., 2015 [40], 6 – Spada et al., 2016 [38].
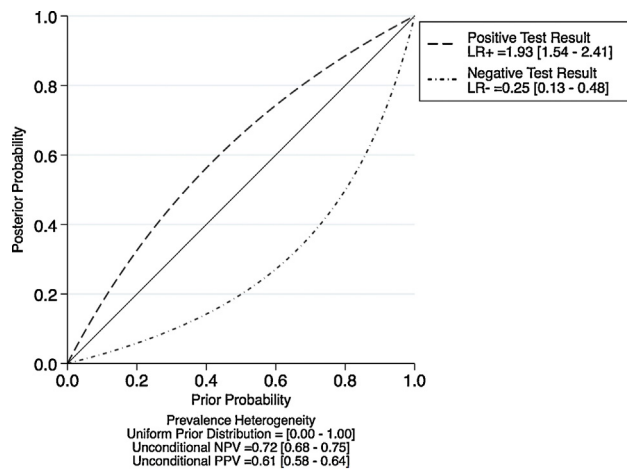

Fig. 3. Meta-analytical probability-modifying plot, illustrating the relationship between the pre-test probability (pre-TP) and post-test probability (post-TP) i.e. psychosis risk in help-seeking subjects following CAARMS assessment, computed as the likelihood of a positive (above diagonal line; LR+) or negative (below diagonal line; LR–) test result over the pre-TP between 0 and 1.

directly test the effect of these differences on the prognostic performance, in the current study, we performed the first meta-analytical comparison of Se and Sp across the CAARMS and SIPS, using previously published SIPS data [14]. We found that Se was higher in the SIPS compared to the CAARMS, while there were no substantial differences in Sp. Overall, it is unlikely that these differences may account for significant differences in the positive predictive values of the two instruments, as confirmed by previous meta-analyses in CHR-P+ samples [18].

In a second step, we estimated the clinical utility of the CAARMS. As previously reported by our lab, clinical utility is not static, instead reliant on the underlying pre-test risk in any given population [49–51]. We found that being classified as CHR+ by the CAARMS is associated with a 16.4% risk of developing psychosis within 2 years, which is lower than the 29.1% 2-year transition risk previously reported [52]. This was driven by a small LR+ (1.9),
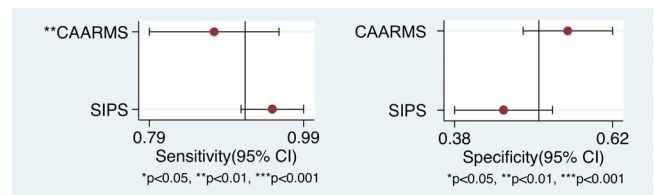

Fig. 4. Meta-regression analyses comparing at meta-analytical level the 2-year sensitivity and specificity of the CAARMS vs. SIPS for the prediction of psychosis. SIPS data taken from [14].

similar to the LR+ seen previously (1.82) [14]. CHR– individuals had 3.38% 2-year transition rate and this was driven by a moderate LR– (0.25), which was not as large as the LR– for CHR assessments as a whole (0.09) [14]. These findings taken altogether indicate that the acceptable prognostic accuracy is due to an imbalance between Se and Sp and LR+ and LR–, with the CAARMS being a valuable tool to correctly identify individuals who will develop psychosis however showing only modest ability to identify those who will not.

On a pragmatic level, the results of this meta-analysis show that the only acceptable prognostic accuracy of the CAARMS needs improving through a refined assessment of psychosis risk. An improved detection of individuals who will transition would lead to improved clinical and research opportunities. For example, a greater proportion of true positives would lead to more efficient primary indicated prevention as well as a more homogenous CHR-P group [48,53] for developing putative treatments. This manuscript has the clinical potential to be the reference point for refining future versions of the CAARMS or for the development of refined prognostic tools and assessments. To improve prediction of psychosis, it seems necessary to tailor it on an individual level. To date, the CAARMS has just considered CHR-P+ individuals as belonging to a whole group. However, it is now clear that such an assumption is incorrect, given the profound difference in level of psychosis risk observed across different CHR-P+ subgroups [48,53]. Furthermore, to date, psychosis prediction has been limited to the assessment and rating of CHR-P symptoms and signs. However, it is evident that these are only epiphenomena of underlying neurobiological and psychological processes that may characterize the onset of psychosis in vulnerable individuals. Research evidence in the field of risk and protective factors associated with an impending vulnerability to psychosis has accumulated over the past few decades and only recently has it been systematically assessed. In a recent large-scale meta-analysis, our lab has stratified the level of evidence for associations of several risk or protective factors and established psychotic disorders [54]. This study may lay the groundwork for investigating how specific risk or protective factors accumulate in CHR-P+ individuals explaining their increased liability to develop psychosis. In a first attempt by our lab [55], we reviewed forty-four studies encompassing 170 independent datasets and 54 risk/protective factors in CHR-P+ individuals. We showed that CHR-P+ individuals were more likely to show obstetric complications, tobacco use, physical inactivity, childhood trauma/emotional abuse/physical neglect, high perceived stress, childhood and adolescent low functioning, affective comorbidities, male gender, single status, unemployment and low educational level as compared to controls. The differential accumulation of these factors in each CHR-P+ individuals are likely to account for the different outcomes observed in these samples, such as psychosis onset, persistence of CHR-P+ features or remission. A refinement of psychosis prediction in these samples would inevitably require a careful investigation of these factors beyond the rating of severity and frequency of CHR-P+ symptoms as currently required by the CAARMS.

4.1. Limitations

Some limitations of this meta-analysis need to be acknowledged. Firstly, only 6 studies were able to be synthesised for this meta-analysis, and although supplying a healthy number of subjects, power could be questioned. Another limitation of our meta-analysis is the small sample size. However, conducting longitudinal studies in individuals assessed for a CHR-P state but not meeting intake criteria is logistically challenging and therefore only a few studies are currently available. Secondly, heterogeneity was very high and this could potentially have been reduced through a greater pool of studies. Thirdly, this heterogeneity remains unexplained as we were unable to perform meta-regressions because there were not enough studies.

5. Conclusion

The 2-year meta-analytical prognostic accuracy of the CAARMS in predicting psychosis is only acceptable. A refined prediction of psychosis risk is necessary to advance clinical research in this area.

Disclosure of interest

The authors declare that they have no competing interest.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.eurpsy.2017.10.001.

References

[1] Insel TR. Rethinking schizophrenia. Nature 2010;468(7321):187–93.
[2] Fusar-Poli P, McGorry P, Kane J. Improving outcomes of first episode psychosis. World Psychiatry 2017;16(3):251–65.
[3] Millan MJ, Andrieux A, Bartzokis G, Cadenhead K, Dazzan P, Fusar-Poli P, et al. Altering the course of schizophrenia: progress and perspectives. Nat Rev Drug Discov 2016;15(7):485–515.
[4] McGorry PD. Early clinical phenotypes, clinical staging, and strategic biomarker research: building blocks for personalized psychiatry. Biol Psychiatry 2013;74(6):394–5.
[5] Fusar-Poli P. Extending the benefits of indicated prevention to improve outcomes of first episode psychosis. JAMA Psychiatry 2017;74(7):667–8.
[6] Fusar-Poli P, Schultze-Lutter F. Predicting the onset of psychosis in patients at clinical high risk: practical guide to probabilistic prognostic reasoning. Evid Based Ment Health 2016;19(1):10–5.
[7] Arbyn M, Verdoodt F, Snijders PJF, Verhoef VMJ, Suonio E, Dillner L, et al. Accuracy of human papillomavirus testing on self-collected versus clinician-collected samples: a meta-analysis. Lancet Oncol 2014;15(2):172–83.
[8] Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. Lancet 2012;379(9833):2279–90.
[9] Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007;335(7611):136.
[10] Fusar-Poli P, Meyer-Lindenberg A. Forty years of structural imaging in psychosis: promises and truth. Acta Psychiatr Scand 2016;134(3):207–24.
[11] Yung AR, Yuen HP, McGorry PD, Phillips LJ, 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(11–12):964–71.
[12] NHS England. Intervention in Psychosis Access and Waiting Time Standard [Internet]. UK: NHS England Publications; 2016 Available from:https://www.england.nhs.uk/mentalhealth/wp-content/uploads/sites/29/2016/04/eip-guidance.pdf.
[13] 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(1):57–66.
[14] Fusar-Poli P, Cappucciati M, Rutigliano G, Schultze-Lutter F, Bonoldi I, Borgwardt S, et al. 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.
[15] Miller TJ, McGlashan TH, Rosen JL, Cadenhead K, Cannon T, Ventura J, et al. Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability. Schizophr Bull 2003;29(4):703–15.
[16] Schultze-Lutter F, Ruhrmann S, Picker H, Klosterkötter J. WC1C development and evaluation of the Schizophrenia Proneness Instrument, Adult version (SPI-A). Schizophr Res 2006;86:S4–5.
[17] Fusar-Poli P, Cappucciati M, Rutigliano G, Lee TY, Beverly Q, Bonoldi I, et al. Towards a standard psychometric diagnostic interview for subjects at ultra high risk of psychosis: CAARMS versus SIPS. Psychiatry J 2016;7146–341.
[18] Fusar-Poli P, Rocchetti M, Sardella A, Avila A, Brandizzi M, Caverzasi E, et al. Disorder, not just state of risk: meta-analysis of functioning and quality of life in people at high risk of psychosis. Br J Psychiatry 2015;207(3):198–206.
[19] Kempton MJ, Bonoldi I, Valmaggia L, McGuire P, Fusar-Poli P. Speed of psychosis progression in people at ultra-high clinical risk. JAMA Psychiatry 2015;72(6):622.
[20] Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283(15):2008–12.
[21] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(7):e1000097.
[22] Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529–36.
[23] Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Handbook for systematic reviews of diagnostic test accuracy. The Cochrane Colloration; 2010.
[24] Smetana GW, Umscheid CA, Chang S, Matchar DB. Methods guide for authors of systematic reviews of medical tests: a collaboration between the Agency for Healthcare Research and Quality (AHRQ) and the Journal of General Internal Medicine. J Gen Intern Med 2012;27(Suppl. 1):S1–3.
[25] Dwamena BA. MIDAS: computational and graphical routines for meta-analytical integration of diagnostic accuracy studies in Stata. Ann Arbor: Division of Nuclear Medicine, Department of Radiology, University of Michigan Medical School; 2007.
[26] StataCorp.. STATA. College Station. TX: StataCorp LP; 2015.
[27] Harbord R, Whiting P. Metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. Stata Journal [Internet] 2009;9(2):211–29. Available from: http://www.stata-journal.com/article.html?article=st0163 [19].
[28] Janda S, Shahidi N, Gin K, Swiston J. Diagnostic accuracy of echocardiography for pulmonary hypertension: a systematic review and meta-analysis. Heart 2011;97(8):612–22.
[29] Hosmer DW, Lemeshow S, May S. Applied survival analysis: regression modeling of time-to-event data. 2nd ed. Hoboken, N.J: Wiley-Interscience; 2008.
[30] Lipsey MW, Wilson DB. Practical meta-analysis. Thousand Oaks, Calif: Sage Publications; 2000.
[31] Bekkar M, Djemaa H, Alitouche T. Evaluation measures for modelsassessment over imbalanced data sets. J Inf Eng Appl 2013;3:27–38.
[32] Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557–60.
[33] Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005;58(9):882–93.
[34] Wanders LK, East JE, Uitentuis SE, Leeflang MMG, Dekker E. Diagnostic performance of narrowed spectrum endoscopy, autofluorescence imaging, and confocal laser endomicroscopy for optical diagnosis of colonic polyps: a meta-analysis. Lancet Oncol 2013;14(13):1337–47.
[35] Li J, Fine JP, Safdar N. Prevalence-dependent diagnostic accuracy measures. Stat Med 2007;26(17):3258–73.
[36] Fusar-Poli P, Rutigliano G, Stahl D, Davies C, De Micheli A, Ramella-Cravaro V, et al. Long-term validity of the At Risk Mental State (ARMS) for predicting psychotic and non-psychotic mental disorders. Eur Psychiatry 2017;42:49–54.
[37] Francesconi M, Minichino A, Carrión RE, Delle Chiaie R, Bevilacqua A, Parisi M, et al. Psychosis prediction in secondary mental health services. A broad, comprehensive approach to the "at risk mental state" syndrome. Eur Psychiatry 2017;40:96–104.
[38] Spada G, Molteni S, Pistone C, Chiappedi M, McGuire P, Fusar-Poli P, et al. Identifying children and adolescents at ultra high risk of psychosis in Italian neuropsychiatry services: a feasibility study. Eur Child Adolesc Psychiatry 2016;25(1):91–106.
[39] Yung AR, Nelson B, Stanford C, Simmons MB, Cosgrave EM, Killackey E, et al. Validation of "prodromal" criteria to detect individuals at ultra high risk of psychosis: 2 year follow-up. Schizophr Res 2008;105(1–3):10–7.
[40] Kotlicka-Antczak M, Pawełczyk T, Rabe-Jabłońska J, Pawełczyk A. PORT (Programme of Recognition and Therapy): the first Polish recognition and treatment programme for patients with an at-risk mental state. Early Interv Psychiatry 2015;9(4):339–42.

[41] Lee J, Rekhi G, Mitter N, Bong YL, Kraus MS, Lam M, et al. The Longitudinal Youth at Risk Study (LYRIKS) – an Asian UHR perspective. Schizophr Res 2013;151(1–3):279–83.

[42] 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(5):894–908.

[43] Addington J, Piskulic D, Perkins D, Woods SW, Liu L, Penn DL. Affect recognition in people at clinical high risk of psychosis. Schizophr Res 2012;140(1–3):Error: FPage (87) is higher than LPage (92)!.

[44] Liu C-C, Lai M-C, Liu C-M, Chiu Y-N, Hsieh MH, Hwang T-J, et al. Follow-up of subjects with suspected pre-psychotic state in Taiwan. Schizophr Res 2011;126(1–3):65–70.

[45] Schultze-Lutter F, Klosterkötter J, Ruhrmann S. Improving the clinical prediction of psychosis by combining ultra-high risk criteria and cognitive basic symptoms. Schizophr Res 2014;154(1–3):100–6.

[46] Simon AE, Grädel M, Cattapan-Ludewig K, Gruber K, Ballinari P, Roth B, et al. Cognitive functioning in at-risk mental states for psychosis and 2-year clinical outcome. Schizophr Res 2012;142(1–3):108–15.

[47] Thomas C, Hyppönen E, Power C. Type 2 diabetes mellitus in midlife estimated from the Cambridge Risk Score and body mass index. Arch Intern Med 2006;166(6):682–8.

[48] Fusar-Poli P, Cappucciati M, Borgwardt S, Woods SW, Addington J, Nelson B, et al. Heterogeneity of psychosis risk within individuals at clinical high risk: a meta-analytical stratification. JAMA Psychiatry 2016;73(2):113–20.

[49] Fusar-Poli P, Rutigliano G, Stahl D, Schmidt A, Ramella-Cravaro V, Hitesh S, et al. Deconstructing pretest risk enrichment to optimize prediction of psychosis in individuals at clinical high risk. JAMA Psychiatry 2016;73(12):1260–7.

[50] Fusar-Poli P, Schultze-Lutter F, Cappucciati M, Rutigliano G, Bonoldi I, Stahl D, et al. The dark side of the moon: meta-analytical impact of recruitment strategies on risk enrichment in the clinical high risk state for psychosis. Schizophr Bull 2016;42(3):732–43.

[51] Fusar-Poli P, Palombini E, Davies C, Oliver D, Bonoldi I, Ramella-Cravaro V, et al. Why transition risk to psychosis is not declining at the OASIS ultra high risk service: the hidden role of stable pretest risk enrichment. Schizophr Res 2017;S0920-9964(17):30353–5.

[52] Fusar-Poli P, Bonoldi I, Yung AR, Borgwardt S, Kempton MJ, Valmaggia L, et al. Predicting psychosis: meta-analysis of transition outcomes in individuals at high clinical risk. Arch Gen Psychiatry 2012;69(3):220–9.

[53] Fusar-Poli P, Cappucciati M, De Micheli A, Rutigliano G, Bonoldi I, Tognin S, et al. Diagnostic and prognostic significance of brief limited intermittent psychotic symptoms (BLIPS) in individuals at ultra high risk. Schizophr Bull 2017;43(1):48–56.

[54] Ramella-Cravaro V, Radua J, Ionnidis JPA, Reichenberg A, Phiphopthatsanee N, Amir T, et al. What causes psychosis? Umbrella review of risk factors. World Psychiatry 2017 (In Press).

[55] Fusar-Poli P, Tantardini M, De Simone S, Ramella-Cravaro V, Oliver D, Kingdon J, et al. Deconstructing vulnerability for psychosis: meta-analysis of environmental risk factors for psychosis in subjects at ultra high-risk. Eur Psychiatry 2017;40:65–75.