

F. Paterlini¹, L. Pelizza¹, G. Galli²,
S. Azzali¹, I. Scazza¹, S. Garlassi¹,
L.R. Chiri³, M. Poletti¹, S. Pupo⁴,
A. Raballo⁵

¹ Department of Mental Health and Pathological Addiction, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy; ² School of Psychiatry, Department of Clinical, Diagnostic and Public Health Medicine, University of Modena and Reggio Emilia, Modena, Italy; ³ Department of Mental Health and Pathological Addiction, Azienda USL di Bologna, Bologna, Italy; ⁴ Intensive Care Unit, Guastalla Civil Hospital, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy; ⁵ Division of Psychiatry, Clinical Psychology and Rehabilitation, Department of Medicine, University of Perugia, Perugia, Italy

Interrater reliability of the authorized Italian version of the Comprehensive Assessment of At-Risk Mental States (CAARMS-ITA)

Summary

Objective

The Comprehensive Assessment of At-Risk Mental States (CAARMS) was specifically developed to detect and assess young individuals at Ultra-High Risk (UHR) of psychosis. Aim of the current study was to test the interrater reliability of the authorized Italian version of the CAARMS (CAARMS-ITA) in young adult help-seekers consecutively recruited through the “Reggio Emilia At-Risk Mental States” (ReARMS) project, an early detection and intervention infrastructure developed in the Reggio Emilia Department of Mental Health.

Methods

We included 51 young adults, aged 18-35 years, seeking help at the Reggio Emilia outpatient mental health services. Two trained raters were paired for each CAARMS interview, both simultaneously in the room with the subject. Interrater reliability of the CAARMS-ITA was tested measuring the Intra-Class Correlation (ICC) coefficients and the Cohen's kappa for interrater agreement on CAARMS-defined diagnosis criteria (i.e. UHR and First-Episode Psychosis [FEP]).

Results

The CAARMS-ITA showed an excellent interrater reliability. The Cohen's kappa for CAARMS diagnoses was 0.845 ($p < 0.001$). The ICC coefficients of the seven CAARMS subscale scores ranged from 0.965 and 0.990.

Conclusions

The CAARMS-ITA is a reliable instrument for detecting and assessing at-risk mental states in Italian clinical setting.

Key words

At-Risk Mental States • Ultra-High Risk • Early detection • Assessment • Prodrome • Psychosis • Schizophrenia

© Copyright by Pacini Editore Srl



OPEN ACCESS

Received: September 16, 2018

Accepted: September 19, 2018

Correspondence

Lorenzo Pelizza
c/o CSM Petrella, via E. Petrella 1/A,
42100 Reggio Emilia, Italy
Tel. +39 0522 339501 • Fax +39 0522 339523
E-mail: lorenzo.pelizza@ausl.re.it

Introduction

Psychosis is a severe psychiatric condition, with schizophrenia being among the main leading causes of disability in young adults in Europe¹. Since the evidence of improvement in patient's functioning is limited once the psychotic disorder is established², early intervention may delay or even avoid First Episode Psychosis (FEP)³.

Within a clinical staging strategy of psychosis, McGorry and colleagues (2003)⁴ proposed the notion of “At-Risk Mental State” (ARMS) to identify individuals at increased risk of developing FEP. The conceptualization of ARMS has to be understood as an early phase of the disease, viewed in perspective, and as the epistemological and nosological reversal of the retrospective concept of prodrome⁵. Within the variety of “ARMS”, the so called “Ultra High Risk” (UHR) criteria were proven to be valid tools to

identify individual with prospectively high (but not inevitable) imminent risk of developing psychosis⁶. Those are: a) Attenuated Psychotic Symptoms (APS), that are experiences of subthreshold positive psychotic symptoms during the past year; b) Brief Limited Intermittent Psychotic Symptoms (BLIPS), which are transient episodes of frank positive psychotic symptoms that have spontaneously remitted within one week, and c) vulnerability, a trait/state risk condition present in individuals who have a first-degree relative with a psychotic disorder or who have a schizotypal personality disorder, along with a significant decrease in functioning during the past year⁷.

Over the years, the UHR criteria have gone through some slight modifications, but the core construct, such as the combination of socio-demographic risk feature (age range: 14-30 years), the state and trait factors (i.e. APS, BLIPS, and vulnerability) and help-seeking behavior, remained the same^{8,9}, with a particular focus on the help-seeking behavior, in order to mitigate the potential high number of false positives that might occur assessing large asymptomatic community samples¹⁰. Several studies tested the predictive strength of UHR criteria to identify individuals at risk of psychosis. The percentage of patients who matched the UHR criteria and subsequently presented a FEP within 12 to 30 months went from 40% in the yearly studies⁶, to 15% in the most recent ones^{11,12}. This decrease in the transition rate was detected also in a recent meta-analysis that provided as a possible explanation of this phenomenon the increased ability in the clinical services to recognize patient at risk and the effectiveness of intervention at an earlier stage¹³.

Among the clinical interviews used to identify UHR/FEP individuals, the “Comprehensive Assessment of At-Risk Mental States” (CAARMS) is one of the most validated and reliable¹⁴. This instrument was explicitly developed at the PACE clinic in Melbourne to assist the early identification, risk stratification and longitudinal monitoring of ARMS⁷. The CAARMS has been adopted, besides Australia, also in many European, Asiatic, and Arabic countries, including UK, France, Spain, Germany, Denmark, Sweden, Greece, Japan, China, Korea, and Tunisia¹⁵.

In this context, the “Reggio Emilia Departmental Group on Early detection and intervention on Psychosis” used the CAARMS in the clinical practice and, through a close collaboration with its Australian authors that granted the copyright in 2008, published the Italian version (CAARMS-ITA) (see supporting information in Pelizza et al., 2018)¹⁶ under the aegis of the “Emilia-Romagna Regional Project on Early Detection in Psychosis”¹⁷.

Aim of the present study was to assess the interrater reliability of the CAARMS-ITA in a sample of Italian young adult help-seekers.

Materials and methods

Participants

The interrater reliability was assessed in 51 young adult (aged 18-35 years) help-seekers consecutively recruited between September 2017 and March 2018 through the “Reggio Emilia At-Risk Mental State” (ReARMS) project, an early detection and intervention infrastructure developed under the governance of the “Regional Project on Early Detection in Psychosis” in the Reggio Emilia Department of Mental Health¹⁸. In the ReARMS project, the participants are assigned to a multidisciplinary team, including a psychiatrist, a clinical psychologist and a case-manager for recovery-oriented early rehabilitation, generally within 2-3 weeks.

ReARMS inclusion criteria were: a) young individuals seeking the help of a specialist; b) age between 13 and 35 years; c) presence of UHR criteria defined by the CAARMS (i.e. APS, BLIPS, and/or Vulnerability), or d) a Duration of Untreated Psychosis (DUP) < 2 years in case FEP is detected in the initial assessment. The exclusion criteria were: a) history of past psychotic episodes either schizophrenic or affective, as specified in the Diagnostic and Statistical manual of Mental Disorders, Fifth Edition (DSM-5)¹⁹; b) history of previous exposure to antipsychotics; c) current substance dependence; d) known mental retardation (Intelligence Quotient < 70); and e) neurological disorders, head injury, or any other medical condition associated with psychiatric symptoms.

All participants entering the ReARMS protocol voluntarily agreed to participate to the study with written informed consent. All individuals assessed in this research were Italian native speakers. Relevant ethical and local NHS research and development approvals were sought for the study.

CAARMS

The CAARMS is a semi-structured clinical interview designed to study different aspects of attenuated psychopathology and functioning (via the integrated Social and Occupational Functioning Assessment Scale [SO-FAS] module)⁷. The administration takes approximately 1-1.5 hours. CAARMS interview is made by 27 items, each one rated (0-6) in terms of intensity and frequency/duration. The items can be clustered in seven subscales: a) “Positive Symptoms” (disorders of thought content, perceptual abnormalities, disorganized speech); b) “Cognitive Change, Attention and Concentration” (subjective experience and observed cognitive change); c) “Emotional Disturbance” (subjective emotional disturbance, observed blunted affect, observed inappropriate affect); d) “Negative Symptoms” (alogia, avolition/apathy, anhedonia); e) “Behavioral Change” (social isolation, impaired role functioning, disorganiz-

ing/odd/stigmatizing behavior, aggressive/dangerous behavior); f) “Motor/Physical Changes” (complaints of impaired motor functioning, impaired bodily sensation, and impaired autonomic functioning); and g) “General Psychopathology” (mania, depression, suicidality and self-harm, mood swings/lability, anxiety, obsessive-compulsive symptoms, dissociative symptoms, impaired tolerance to normal stress).

Among those subscales the “Positive Symptoms” one is used to determine both the UHR criteria and the threshold for psychosis, i.e. the presence of fully (positive) psychotic symptoms occurring for at least 1 week (either on a daily basis or more than three times a week) with each symptom continuing for more than 1 hour on each occasion⁷. UHR status is defined as follows: a) vulnerability group: schizotypal personality disorder in the individual or family history of psychosis in a first-degree relative combined with 30% drop in functioning for at least 1 month or chronic low functioning, as measured by the SOFAS (the decline in functioning is calculated by subtracting the current SOFAS score from the highest SOFAS score in the last year; scores range from 1 to 100); 2) APS group: sub-threshold positive psychotic symptoms within the past 12 months; and 3) BLIPS group: criteria for psychosis met for less than 7 day at a time and ceasing spontaneously (i.e. without antipsychotic medication)⁷.

The Australian version of the CAARMS was translated into Italian by Andrea Raballo and back-checked by a team of experienced mental health professionals after obtaining permission from the original authors. This early version was then examined and judged as satisfactory by a staff member of the PACE clinic in Melbourne, who was fluent in Italian and familiar with the usage of the CAARMS¹⁶.

Procedures

The interrater reliability of the CAARMS-ITA was tested by using data from consecutive joint interviews of 51 young adults entering the ReARMS protocol. Initially, three psychologists of the ReARMS project with clinical experience of psychotic disorders were trained on the usage of the CAARMS through collective supervision by the main author of the approved Italian translation¹⁷, who was trained at Orygen, The National Centre of Youth Mental Health in Melbourne, Australia. Preliminary administration of the instrument to suspected ARMS individuals was conducted before the study. Of the three raters, two were paired for each interview, both simultaneously in the room with the subject. Interrater agreement was also assessed for the UHR/FEP criteria.

Statistical analysis

The interrater reliability is a method established to test the agreement among the various data collectors and

it measures the extent to which raters assign the same score to the same variable. This method allows to know which extent the data collected in the study are correct representations of the variables measured²⁰. To compute the data for the interrater reliability of the CAARMS-ITA, we used two different statistic tools: the Intra-Class Correlation (ICC) coefficients and the Cohen's kappa.

To assess the interrater reliability of the CAARMS-ITA subscale scores, we used the two-way, mixed effect, model of ICC, a tool commonly applied for ordinal, interval, and ratio variables²¹. In the current study, we also focused on the absolute agreement. Moreover, to generalize the reliability of multiple raters to the subjects rated by one coder, the ICC method here used was for single-measures²¹.

The Cohen's kappa is used for a set of nominal ratings to measure the observed level of agreement between coders and allows to correct agreement that would be expected by chance²¹. In our study, the nominal variables were the three CAARMS-defined diagnoses: a) UHR- (i.e. participants below the UHR threshold to be considered at risk of developing psychosis); b) UHR+ (i.e. participants who met the UHR criteria), and c) FEP (participants who met the FEP criteria).

Data analysis was performed using the software Statistical Package for Social Science (SPSS) 15.0 for Windows²².

Results

Sample

The interrater reliability of the CAARMS-ITA was assessed on 51 ReARMS young adult participants. Of these subjects, 26 (51%) were males. After distributions of all quantitative variables were examined for violations of normality assumption, descriptive analyses included median and interquartile range for not normally distributed parameters. The median was 25 years (interquartile range = 21-32 years) for age, 13 years (interquartile range = 10-15 years) for education, and 50 weeks (interquartile range = 15-56 weeks) for the Duration of Untreated Illness (DUI), defined as the interval (in weeks) between the onset of a psychiatric clinically relevant symptom and the administration of the first pharmacological/psychological treatment²³ (Tab. I).

Interrater reliability

The Cohen's Kappa for the three CAARMS diagnoses (i.e. UHR-, UHR+, and FEP) was 0.845 ($p < 0.001$), showing an almost perfect agreement between raters²⁴. Similarly, the result of the overall ICC coefficient was 0.990, demonstrating an excellent interrater reliability. The analysis was also performed on the seven CAARMS subscores, on each item of the interview, and

TABLE I. Sociodemographic and clinical characteristics of the total sample ($n = 51$).

Variables	
Gender (males)	26 (51%)
Age	25 (21-32)
Education (in years)	13 (10-15)
DUI (in weeks)	50 (15-56)

Note. DUI: Duration of Untreated Illness. Frequency (and percentage), median and interquartile range are reported.

on the SOFAS score. Interrater reliability ranged from very good to excellent for all the scores. The results of the individual ICC coefficients are reported in Table II.

Discussion

The assessment of inter-rater reliability provides a way to quantify the level of agreement between two or more coders who make independent ratings on a variable. This analysis is necessary for research studies where data are collected through ratings provided by different coders. The presence of a good interrater reliability is one of the more relevant factor demonstrating the general reliability of an interview²⁰. In the current study, reliability of the Italian version of the CAARMS was assessed with respect to interrater reliability. The overall ICC coefficient of the CAARMS-ITA was 0.990 and the coefficients for each subscale showed good to excellent reliability, in line with the original validation study by Yung et al. (2005)⁷. Moreover, in an Italian sample of 34 UHR+ young adults, the interrater reliability of the CAARMS had been previously assessed in a pilot study by Fusar-Poli et al. (2012)²⁵ using an unofficial and non-authorized version of the interview, which showed ICC scores comparable to ours. Finally, the CAARMS has been also validated in other different languages, such as Japanese²⁶, Greek²⁷, and Arabic²⁸, with an interrater reliability ranging from good to excellent.

Overall, these findings suggest that the interrater reliability of the CAARMS-defined UHR criteria were satisfactory, and that this instrument can be safely administered by trainer raters in clinical and research settings to assess the broad spectrum of prodromal and psychotic symptoms presented by young help-seekers referred to mental health services.

Limitations

There are several methodological limitations for the current study. Firstly, the group of raters was small and had considerable clinical experience with prodromal symptoms of psychosis. This can prevent the generalizability of our findings to primary care setting with no or less

TABLE II. Intra-Class Correlation (ICC) coefficients of the CAARMS-ITA.

CAARMS subscale	ICC
Overall	0.990
1. Positive symptoms	0.990
1.1 Unusual thought content	0.988
1.2 Non-bizarre ideas	0.978
1.3 Perceptual abnormalities	0.977
1.4 Disorganized speech	0.986
2. Cognitive change	0.973
2.1 Subjective cognitive change	0.955
2.2 Objective cognitive change	0.967
3. Emotional disturbance	0.987
3.1 Subjective emotional disturbance	0.952
3.2 Blunted affect	0.988
3.3 Inadequate affect	0.981
4. Negative symptoms	0.979
4.1 Alogia	0.947
4.2 Avolition/apathy	0.961
4.3 Anhedonia	0.948
5. Behavioral change	0.982
5.1 Social isolation	0.967
5.2 Impaired role functioning	0.936
5.3 Disorganized behavior	0.980
5.4 Aggressive behavior	0.968
6. Motor/physical change	0.966
6.1 Subjective motor functioning	0.997
6.2 Objective motor functioning	0.990
6.3 Subjective bodily sensation	0.939
6.4 Subjective autonomic functioning	0.923
7. General psychopathology	0.965
7.1 Mania	0.974
7.2 Depression	0.924
7.3 Suicidality/self-harm	0.957
7.4 Affective instability	0.956
7.5 Anxiety	0.937
7.6 OCD	0.985
7.7 Dissociative symptoms	0.965
7.8 Impaired subjective tolerance to normal stress	0.931
SOFAS score	0.974

Note. ICC: Intra-Class Correlation coefficients; CAARMS-ITA: the authorized Italian version of the Comprehensive Assessment of At-Risk Mental States; OCD: Obsessive-Compulsive Disorder; SOFAS: Social and Occupational Functioning Assessment Scale.

experience in the UHR assessment. Secondly, UHR+ sample size was limited. Thus, interrater reliability of the CAARMS-ITA must be studied in a larger UHR+ sample. Finally, our UHR+ participants were referred to the ReARMS program because considered potentially at risk for psychosis and thus the results might not be generalizable to help-seeking population in general mental health services.

Conclusions

Despite these limitations, CAARMS-ITA demonstrated to have an excellent interrater reliability and to be a reliable tool to assess and detect ARMS in Italian clinical setting.

References

- 1 Schmidt SJ, Schultze-Lutter F, Schimmelmann BG, et al. *EPA guidance on the early intervention in clinical high risk states of psychoses*. *Eur Psychiatry* 2015;30:388-404.
- 2 Oliver M, Kotlicka-Antczak A, Minichino G, et al. *Meta-analytical prognostic accuracy of the Comprehensive Assessment of at Risk Mental States (CAARMS): the need for a refined prediction*. *Eur Psychiatry* 2018;49:62-8.
- 3 Azzali S, Pelizza L, Paterlini F, et al. *Reliability of the Italian version of the 16-item Prodromal Questionnaire (iPQ-16) for psychosis screening in a young help-seeking community sample*. *Journal of Psychopathology* 2018;24:16-23.
- 4 McGorry PD, Yung AR, Phillips LJ. *The "close-in" or ultra-high risk model: a safe and effective strategy for research and clinical intervention in pre-psychotic mental disorder*. *Schizophr Bull* 2003;29:771-90.
- 5 Pelizza L, Raballo A, Semrov E, et al. *Identification of young people at Ultra-High Risk (UHR) of developing psychosis: validation of the "Checklist per la valutazione dell'esordio psicotico" for use in primary care setting*. *Journal of psychopathology* 2016;22:172-9.
- 6 Yung AR, Phillips LJ, Yuen HP et al. *Psychosis prediction: 12-month follow-up of a high-risk ("prodromal") group*. *Schizophr Res* 2003;60:21-32.
- 7 Yung AR, Yuen HP, McGorry PD et al. *Mapping the onset of psychosis: the comprehensive assessment of at-risk mental states*. *Aust N Z J Psychiatry* 2005;39:964-71.
- 8 Yung AR, Nelson B, Stanford C, et al. *Validation of "prodromal" criteria to detect individuals at ultra-high risk of psychosis: 2-year follow-up*. *Schizophr Res* 2008;105:10-7.
- 9 Raballo A, Larøi F. *Clinical staging: a new scenario for the treatment of psychosis*. *Lancet* 2009;374:365-7.
- 10 Ramella Cravaro V, Raballo A. *Early detection of schizophrenia: a clinical-psychopathological revision of ultra-high risk approach*. *Journal of Psychopathology* 2014;20:442-50.
- 11 Nelson B, Yuen HP, Wood SJ, et al. *Long-term follow-up of a group at ultra-high risk ("prodromal") for psychosis: the PACE 400 study*. *JAMA Psychiatry* 2013;70:793-802.
- 12 Hartmann JA, Yuen HP, McGorry PD, et al. *Declining transition rates to psychosis disorder in "ultra-high risk" clients: investigation of a dilution effect*. *Schizophr Res* 2016;170:130-6.
- 13 Fusar-Poli P, Bonoldi I, Yung AR, et al. *Predicting psychosis: meta-analysis of transition outcomes in individuals at high clinical risk*. *Arch Gen Psychiatry* 2012;69:220-9.
- 14 Pelizza L, Raballo A, Semrov E, et al. *Validation of the early detection Primary Care Checklist in an Italian community help-seeking sample: the "Checklist per la valutazione dell'esordio psicotico"*. *Early Interv Psychiatry* 2017;July 26. [Epub head of print]
- 15 Braham A, Bannour AS, Romdhane AB, et al. *Validation of the Arabic version of the Comprehensive Assessment of At Risk Mental States (CAARMS) in Tunisian adolescents and young adults*. *Early Interv Psychiatry* 2014;8:147-54.
- 16 Pelizza L, Paterlini F, Azzali S, et al. *The approved Italian version of the comprehensive assessment of at-risk mental states (CAARMS-ITA): field-test and psychometric features*. *Early Interv Psychiatry* 2018 Apr 26. [Epub ahead of print]
- 17 Raballo A, Semrov E, Bonner Y, et al. *Traduzione e adattamento italiano della CAARMS (the Comprehensive Assessment of At Risk Mental States)*. Bologna: Regione Emilia-Romagna 2013.
- 18 Raballo A, Chiri LR, Pelizza L, et al. *Field-testing the early intervention paradigm in Emilia-Romagna: the Reggio Emilia At Risk Mental State (ReARMS) Project*. *Early Interv Psychiatry* 2014;8:88.
- 19 American Psychiatric Association (APA). *Diagnostic and Statistical Manual of mental disorders, V ed*. Arlington: American Psychiatric Publishing Inc 2013.
- 20 McHugh ML. *Interrater reliability: the kappa statistic*. *Biochem Med (Zagreb)* 2012;22:276-82.
- 21 Hallgren MA. *Computing inter-rater reliability for observational data: an overview and tutorial*. *Tutor Quant Methods Psychol* 2012;8:23-34.
- 22 SPSS Inc. *Statistical Package for Social Science (SPSS) 15.0 for Windows*. Chicago, IL: SPSS Press 2010.
- 23 Rapp C, Canela C, Studerus E, et al. *Duration of untreated psychosis/illness and brain volume changes in early psychosis*. *Psychiatry Res* 2017;255:332-7.
- 24 Landis JR, Koch GG. *The measurement of observer agreement for categorical data*. *Biometrics* 1977;33:159-74.
- 25 Fusar-Poli P, Hobson R, Raduelli M, et al. *Reliability and validity of the Comprehensive Assessment of the At Risk Mental States, Italian version (CAARMS-I)*. *Curr Pharm Des* 2012;18:386-91.
- 26 Miyakoshi T, Matsumoto K, Ito F, et al. *Application of the Comprehensive Assessment of At-Risk Mental States (CAARMS) to the Japanese population: reliability and validity of the Japanese version of the CAARMS*. *Early Interv Psychiatry* 2009;3:123-30.
- 27 Kollias C, Kontaxakis V, Havaki-Kontaxaki B, et al. *Interrater reliability of the Greek version of CAARMS among two groups of mental health professionals*. *Psychiatriki* 2015;26:217-22.