

Speech disturbances analysis in ultra-high risk of psychosis (UHR): a systematic review.

Ignacio Olivares R.^{1,2} Alicia Figueroa B.^{2,3,4}

Introduction: The need for early intervention in psychosis has led to searching for useful biomarkers in its prediction, where the analysis of spoken language stands out for its easy obtaining and low economic cost. In this systematic review, we analyze the main speech disturbances in patients at ultra-high risk for psychosis (UHR), evaluating their prevalence and their relationship with transition to psychosis. **Methods:** A search was carried out in PubMed and Embase databases for studies in English or Spanish, as well as the reference lists of the articles found. **Results:** Of 140 articles identified, 15 were included. The variables of the language analyzed were Linguistic Cohesion, Semantic Coherence, Syntactic Complexity, Metaphorical Production, Emotional Prosody and Formal Thought Disorder (FTD). The prevalence found ranged between 21% and 85% for disturbances included within the FTD, with not prevalence measures for the rest of the variables. The global accuracy for UHR transition to psychosis ranged from 70% to 100% across studies. **Conclusion:** There is a wide range of speech disorders in UHR patients, where the FTD is the most studied one. The prevalence of these alterations seems to be high, especially with regard to FTD. The analysis of those speech alterations in UHR patients appears as an excellent tool to predict transition to psychosis, particularly through interview transcription and the use of Artificial Intelligence.

Keywords: Ultra-high risk for psychosis; Psychosis; Speech.

1. INTRODUCTION

Language disturbances have been deeply studied in mental pathologies, such as major depressive disorder, mania and specially schizophrenia⁽¹⁾⁽²⁾. Currently language is deemed as a predictive biomarker of psychosis⁽³⁾⁽⁴⁾. One of the main sources of language studying schizophrenia is the so called Formal Thought Disorder (FTD), which comprises various

symptoms centrally manifested in psychotic disorders, whose presence has been correlated with poor functioning and evolution of the disease⁽⁵⁾⁽⁶⁾. FTD Clinical Evaluation, was proposed by Andreasen, in 1979 with the Scale named Thought, Language and Communication (TLC) who reformulated it, along with Grove, in 1986⁽⁷⁾⁽⁸⁾. TLC gathers language and thought disturbances arising in FTD, in Subscales (positive, positive objective, negative, disorgani-

The authors declare that they have no conflicts of interest with respect to this article.

Accepted: 2021/08/18

Received: 2020/12/16

¹ Medicine student, Universidad de Chile, Faculty of Medicine.

² Psiquislab Laboratory of Translational Psychiatry and Psychopathology.

³ ANID, Milenio Scientific Initiative Program, Núcleo Milenio Imhay.

⁴ PhD in Clinical Linguistics, Universidad de Chile, Faculty of Medicine, Psychiatry & Mental Health Department, Campus Sur.

zed), and in turn in 18 factors ranged from (0) absent or not observed to (4) severe. However, in the literature⁽⁹⁾ these divisions vary, according to subscales.

During FTD exploration and in other language disturbances, as psychosis biomarkers, various applications for Artificial Intelligence have been developed. By means of training non supervised learning models, better known as Natural Language Processing (NLP), successful automatic identification of some psychosis predicting characteristics in population at risk has been achieved, thus becoming a promising tool for supporting early diagnosis of severe mental diseases⁽¹⁰⁾⁽¹¹⁾.

During the last decades the scientific community has focused on identifying predictors at the prodromal phase of psychosis, where some disturbances in psychosocial areas may appear, along with nonspecific symptomatology⁽¹²⁾. This stage becomes an opportunity for early intervention and reduction of the transition rate to psychosis⁽¹³⁾. This has led to the development of specific criteria aimed to identify people with high-risk clinical characteristics, for further development of psychosis, known as High risk mental condition HMRC. Various instruments for determining EMAR have been designed, among them we can find semi structured psychometric interviews, such as CAARMS (Comprehensive Assessment of At-Risk Mental States)⁽¹⁴⁾ and SIPS/SOPS (Structured Interview for Prodromal Symptoms/Scale of Prodromal Symptomatology)⁽¹⁵⁾. Both instruments have proved to have high sensitivity to detect EMAR. However, these have a low specificity, which implies that only 29% of EMAR patients transitioned to psychosis within 2 years, reaching up to 36%, after 3 years of follow-up⁽¹⁶⁾⁽¹⁷⁾.

The main objective of this systematic review is to propose an analysis of studies including spoken language disturbances in EMAR population. Likewise, the specific objectives are: 1) to identify the main speech disturbances in EMAR patients, 2) to determine prevalence of speech disturbances and 3) to evaluate the correlation among speech disturbances and transition to psychosis.

2. METHODS

2.1 Sources of information

This systematic review was carried out by following PRISMA Guidelines, by searching in PubMed and Embase, on July 12th, 2020, with no restrictions in year of publication. Additionally, manual search is performed in the reference lists of the articles found, in order to identify relevant publications.

2.2 Search Strategy

Key Words and the specific strategy for each data base is further detailed in Table 1.

2.3 Eligibility Criteria

Inclusion criteria for the studies were as follows: 1) cross sectional design (for Prevalence Rates) or longitudinal (for the rates of transition to psychosis), 2) EMAR diagnosis made by using one of the following instruments: Comprehensive Assessment of at Risk Mental States (CAARMS)⁽¹⁴⁾, or Structured Interview for Prodromal Syndromes/Scale of Prodromal Symptoms (SIPS/SOPS)⁽¹⁵⁾, 3). These are instruments for diagnosing transition to psychosis and for analyzing clearly specified language disturbances, and 4) Studies in English or in Spanish.

Only studies measuring speech variables will be included. Analysis of written language or gestures will not be considered. Additionally, only studies where open answers will be chosen are evaluated, excluding multiple choice tasks, with a limited number of potential answers.

Studies were not restricted per age of the participants, duration of the language evaluation strategy, follow-up duration (in longitudinal studies), or according to relation with antipsychotic treatment.

2.4 Data Retrieval

For each paper, the following data key were assessed: 1) study design; 2) sample characteristics; 3) EMAR diagnosis criteria and transition to psychosis; 4) characteristics of the language evaluation strategy; 5) language variables analyzed; 6) main results and statistics accuracy; and 7) software used (if available).

2.5 Recorded Variables

I. Oral Language: Specifically dealing

Table 1: Specific Search Strategies

Embase	<ol style="list-style-type: none"> 1. 'ultra high risk':ab,ti OR 'ultra-high risk':ab,ti OR 'ultra high risk for psychosis':ab,ti OR 'ultra high risk of psychosis':ab,ti OR 'UHR':ab,ti OR 'at risk mental state':ab,ti OR 'at-risk mental state':ab,ti OR 'at risk mental state for psychosis':ab,ti OR 'ARMS':ab,ti OR 'clinical high risk':ab,ti OR 'clinical high risk for psychosis':ab,ti OR 'clinically high risk for psychosis':ab,ti OR 'CHR':ab,ti OR 'psychotic prodrome':ab,ti 2. 'language':ab,ti OR 'linguistics':ab,ti OR 'speech':ab,ti 3. 'psychosis':ab,ti OR 'schizophrenia':ab,ti 4. #1 AND #2 AND #3
PubMed	<ol style="list-style-type: none"> 1. ultra high risk[Title/Abstract] OR ultra-high risk[Title/Abstract] OR ultra high risk for psychosis[Title/Abstract] OR ultra high risk of psychosis[Title/Abstract] OR UHR[Title/Abstract] OR at risk mental state[Title/Abstract] OR at-risk mental state[Title/Abstract] OR at risk mental state for psychosis[Title/Abstract] OR ARMS[Title/Abstract] OR clinical high risk[Title/Abstract] OR clinical high-risk[Title/Abstract] OR clinical high risk for psychosis[Title/Abstract] OR clinically high risk for psychosis[Title/Abstract] OR CHR[Title/Abstract] OR psychotic prodrome[Title/Abstract] 2. language[Title/Abstract] OR linguistics[Title/Abstract] OR speech[Title/Abstract] 3. psychosis[Title/Abstract] OR schizophrenia[Title/Abstract] 4. #1 AND #2 AND #3

with language structure disturbances and/or discursive content regarding in teraction.

- II. EMAR: Criteria to define EMAR patients include 3 symptomatic groups: a) “Attenuated Psychotic Symptoms” (APS) with positive psychotic symptoms with an intensity or under-threshold frequency, b) “Brief Limited and Intermittent Psychotic symptoms” (BLIPS) with brief psychotic episodes, less than 1 week of duration, which remits spontaneously with no need to use antipsychotic drugs, and c) “Genetic risk” (GRD) including people with schizotypal personality disorder and/or who have a first-grade relative with some psychotic disorders, combined with a significant decrease in the functioning level.
- III. Instruments for EMAR diagnosis: CAARMS and SIPS/SOPS are psychometric interviews validated and spread all over the world⁽¹⁴⁾⁽¹⁵⁾. Both contain various sections comprising positive, negative, cognitive, emotional/affective symptoms, among others.
- IV. Prevalence and correlation with transition: Both measures were obtained by statistical analysis of the sample.

3. RESULTS

3.1 Studies Selection

The flow chart of the studies selected in the

systematic review is described in the Figure #1. In total, 235 records were found (225 from the search made in PubMed and Embase, plus 10 crossed references from the identified articles). Eligibility of 35 full text articles was evaluated. From all of them, 20 did not meet the criteria, therefore, they were excluded from the qualitative analysis. As a result, the studies included in this systematic review turned out to be 15.

3.2 Studies Characteristics

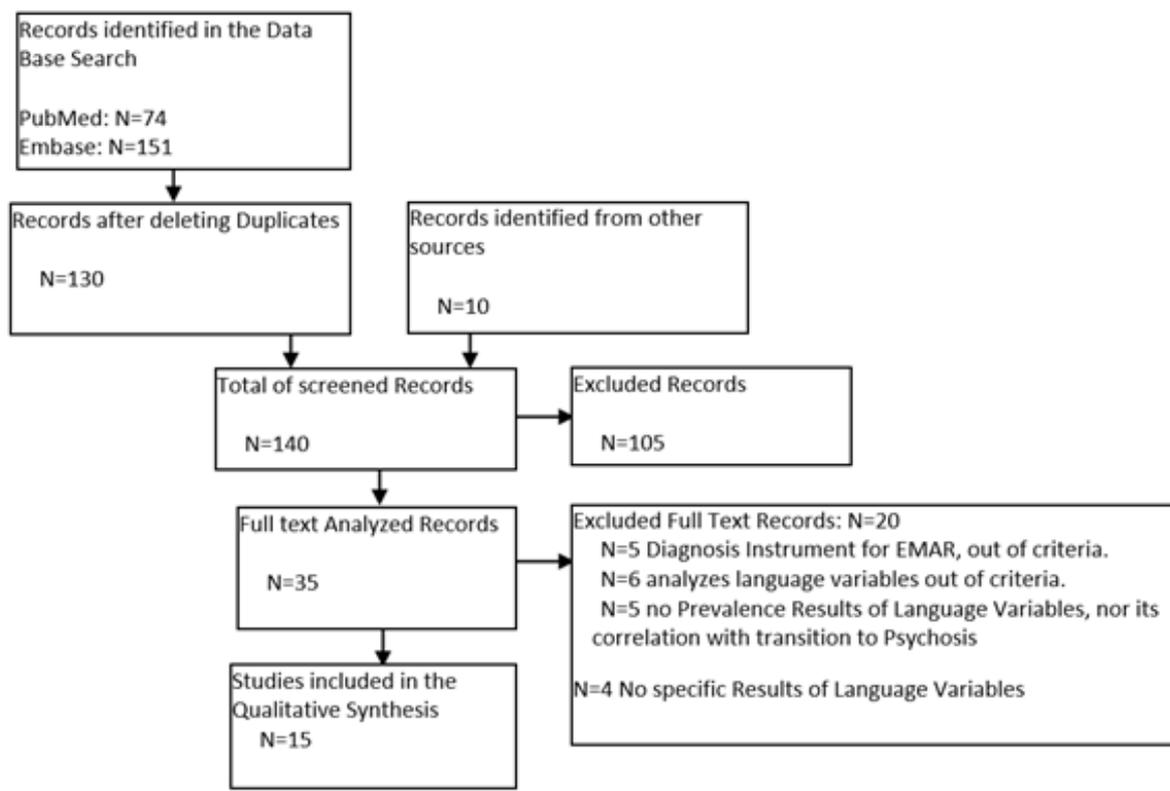
The list of studies included and their main characteristics are described in the Table 2.

All of these studies had a longitudinal design; 10 of them were prospective; 5 of them were retrospective, using transcribed interviews or scores obtained from other studies. 11 of them used SIPS/SOPS and 4 CAARMS for EMAR diagnosis. The size of the samples analyzed varied between 10 to 744 EMAR patients.

For the diagnosis of transition to psychosis (EMAR+) various criteria were used. SIPS/SOPS, CAARMS, DSM-IV, SCID-I, BPRS/CASH and CIE-10, alone or in combination.

Results of speech variables were obtained from 1) records and tasks transcription, open oral interviews, semi structured interviews or from induced free speeches, or 2) ranking of semi structured interviews. 60%⁽⁹⁾ studies used this second methodology, as a language evaluation strategy, by analyzing the score of one specific item from CAARMS or SIPS/SOPS. CAARMS “Disorganized language” is an item

Figure 1. Flowchart for Studies Selection



aimed to state an approximate valuation of speech performance. In a very general manner, lexical-semantic adequacy of the answers is evaluated, according to the interviewee skills for devising ideas and omit irrelevant information for the communication context⁽¹⁴⁾. On its part, the item “Disorganized Communication” included in SIPS/SOPS provides a general evaluation of the verbal communication and speech coherence during the interview. It reviews difficulties for keeping communication objectives, observing adequacy level of linguistic structures used⁽¹⁵⁾. Both items are focused on detecting FTD –related disturbances, such as tangentiality, content poorness, incoherence, among others. For full development of the items, both subjective appreciation questions made to the interviewee, as well as an objective evaluation made by the interviewer are required, thus resulting in a final score with a significant subjective component. This situation differentiates the studies that used these items as a Language Evaluation Method of the remaining 40% of studies, which performed a direct analysis of the transcribed answers from EMAR patients.

Additionally, the strategy duration ranking items is undetermined, as the analysis is performed during the whole interview, and there is no standardized time for such. That is why we have decided to present the results in 2 sections, in the Table 2, studies using strategies for evaluating language by means of transcribed interviews (Section A) and studies using as a strategy the score analysis of an item of its diagnosis instrument for EMAR (Section B).

Duration of the strategies in Section A ranged from 8-minute structured tasks to narrative open interviews ~60 minutes.

LANGUAGE VARIABLES

Section A: Linguistic Cohesion, Semantic Coherence, Syntactic Complexity, Metaphoric Production, Emotional Prosody and FTD variables were analyzed.

Section B: An item of CAARMS or SIPS/SOPS was used. Both items are FTD indicators and show Content Poorness, Loss of Associations, Tangentiality, Circumstantiality and/or Incoherence along the interview.

Prevalence of Language Disturbances

Section A: A study reported the FTD Prevalence levels in EMAR, according to the TLI (Thought and Language Index). It is an index aimed to evaluate 8 speech abnormalities, in response to the Rorschach's Test or to pictures in the Thematic Perception Test. For the Subscale "TLI Positive/Disorganized" it was 75%; and for the "TLI Negative" Subscale it was 21% for scores ≥ 0.5 in a scale ranging from 0.25 (questionable disturbance) to 1 (clearly altered) according to its severity. Within the Subscale "TLI Positive/Disorganized" the items "Use of Peculiar Sentences" and "Use of Peculiar Words" had a Prevalence of 72% and 25% respectively, for scores ≥ 0.5 in severity.

Section B: Four studies reported FTD Prevalence in EMAR; 26.6% for scores ≥ 4 in severity or ≥ 3 in frequency of CAARMS and 29.7%, 37.1% and 44% for scores ≥ 3 , in SIPS/SOPS severity.

Transition to Psychosis and Language Disturbances

For the statistical analysis of the correlation with the transition were applied various statistical models; logistic regression (with its results, such as OR Odds Ratio), Cox's regression (with its results, such as HR Hazard Ratio) and convex envelope (convex hull, with its results, such as global accuracy percentage).

Section A: Two studies obtained mixed results. Some items were correlated with the transition and other were not. Among those not related, the full LTI score, the "TLI Negative" Subscale and the "TLI Positive/Disorganized" Subscale are reported. On the other hand, the items "Speech Poorness" and "Frequency of use of Determinants" were not a statistically significant transition predictor ($p > 0.05$).

In all the studies of this section, at least, one transition predictor item was found. In a study, the frequency of the item "Illogical Thought" was a statistically significant transition predictor ($p = 0.023$, OR=4.64; 95% IC=1.24-17.41). For the remaining studies, global accuracies were reported between 70% to 100% for the transition to psychosis. Including within the analysis, 2 combined items as predictors, up to a set of 5 items.

Section B: statistically significant transition

predictors were found ($p < 0.05$) for the FTD variable of the SIPS/SOPS and the CAARMS, with various results, according to the model of analysis and the cutting score, for the variable severity. When scores ≥ 3 and ≤ 5 were considered, an HR=8.9; 95% IC=2.49-42.41 was obtained. For scores ≥ 3 an HR=2.63; 95% IC=1.18-5.85 was obtained. Four studies analyzed the same variable, but with no minimum severity score, thus obtaining an OR=1.43; 95% IC=1.04-1.97 and OR=2.1; 95% IC=1.03-4.27. Additionally, an HR=1.27; 95% IC=1.08-1.50 and HR=1.69; 95% IC=1-39-2.05.

Additionally, a study classified severity scores of the FTD variable in 4 paths, according to their evolution within 30 months of follow-up. The Path "Persistently High" of the variable was reported to be a statistically significant transition predictor ($p < 0.05$, HR=2.23; 95% IC=1.01-4.93).

Severity of the item "Disorganized Communication /Conceptual Disorganization" was not a statistically significant transition predictor ($p > 0.05$) for 2 studies.

4. DISCUSSION

4.1 Abstract

Use of AI to perform an automated language analysis in psychiatric pathologies is an innovating tool for universal clinical application, as may be seen with this systematic review. Results state that when searching for language anomalies in EMAR population, some variables are more specific for evaluating speech and others are less sharp. The analysis was made with various instruments or softwares and by using various methodologies. The language variables searched are mainly, Linguistic Cohesion, Semantic Coherence, Syntactic Complexity, Metaphoric Production, Emotional Prosody and FTD. According to the analysis, a Prevalence from 21% up to the 85% was reported for disturbances included within FTD. There are no prevalence measures for the remaining variables.

In studies using the score of an item for the research, speech disturbances were found or items which, by means of the retrospective analysis of the samples, significantly predicted transition to psychosis, by using logistic

regression measures and Cox's Regression.

For those studies using transcriptions for their research, in less than 60 minutes these could obtain an amount of speech enough as to perform transition predictions with an accuracy ranging from 70% to 100%. Even more, some authors estimate that application and measurement of the CAARMS item "Disorganized language" (unspecific measurement) could be made in less than 10-15 minutes, while for applying full CAARMS nearly 120 minutes are required⁽⁴⁷⁾. These characteristics could assign to language analysis a significant role in EMAR patients follow-up.

4.2 Limitations

FTD analysis, based on the items "Disorganized language" and "Disorganized Communication" of EMARS diagnosis interviews have a significant subjective component, which makes them partly hard to use in statistical analysis. For instance, one of the questions included in CAARMS subjective evaluation is: "¿Do you use words that are not relevant, or are totally impertinent?"⁽¹³⁾. It is important to highlight that the answer to this question mostly depends on the context, both the patient's context and the interviewer's context. In order to mitigate this effect, in the studies inter-interviewer variability scores are usually obtained, which allow to objectify criteria, in a certain manner, when interviews are ranked, as additionally, no accurate limits exist to determine each value within the scores (Example: 0-6 in CAARMS, 0.25-1 in TLI).

An EMAR subject may remain in this phase of under-threshold symptoms, to make his/her transition to psychosis or have remission from his/her condition⁽⁴⁸⁾. It follows from this that clinical characteristics could evolve similarly; therefore, prevalence measurements of language disturbances will depend on the time the subject is evaluated by Health Professionals, on the follow-up duration; and on the frequency of controls. We could expect higher prevalence rates in studies with a longer duration. On the other hand, the statistical analysis performed by the studies, in order to determine the predictive power of a language variable regarding transition are diverse and consider various mathematic models, which can make general interpreta-

tion more complex.

4.3 CONCLUSIONS

There is a wide range of speech disturbances in EMAR patients' speech. The most studied disturbance is Formal Thought Disorder (FTD). Prevalence of language disturbances in EMAR, especially regarding FTD, seem to be high, although it is highlighted when these are measured all together. Language disturbances analysis, specially made by means of interviews transcription and computing methods seems to be an excellent tool for predicting transition to psychosis.

It is necessary to keep developing instruments for specifically measuring language disturbances in EMAR population, and also to perform studies focused on language analysis in other foreign languages, other than English, as these offer a cost-effective alternative, are easy to obtain and provide promising results, plus having a great potential to be used in clinical internship.

5. REFERENCES

1. Kuperberg GR. Language in schizophrenia part 1: an introduction. *Lang Linguist Compass*. 2010;4:576-589.
2. Docherty NM. Cognitive Impairments and Disordered Speech in Schizophrenia : Thought Disorder , Disorganization , and Communication Failure Perspectives. *J Abnorm Psychol*. 2005;114:269-278.
3. de Boer JN, Brederoo SG, Voppel AE, Sommer IEC. Anomalies in language as a biomarker for schizophrenia. *Curr Opin Psychiatry*. 2020;33(3):212-218.
4. Argolo F, Magnavita G, Mota NB, Ziebold C, Mabunda D, Pan PM, et al. Lowering costs for large-scale screening in psychosis: a systematic review and meta-analysis of performance and value of information for speech-based psychiatric evaluation. *Braz J Psychiatry*. 2020;42(6):673-686.
5. Bleuler E. (Trans. J. Zinkin), Zinkin J. *Dementia Praecox or the Group of Schizophrenias*. International Universities Press; New York: 1950.
6. Norman RM, Malla AK, Cortese L, Cheng

- S, Diaz K, McIntosh E, et al. Symptoms and cognition as predictors of community functioning: a prospective analysis. *Am J Psychiatry* 1999; 156: 400-5.
7. Andreasen NC. Thought, language, and communication disorders. I. Clinical assessment, definition of terms, and evaluation of their reliability. *Arch Gen Psychiatry*. 1979;36(12):1315-1321.
 8. Kircher, T., Krug, A., Stratmann, M., Ghazi, S., Schales, C., Frauenheim, et al. (2014). A rating scale for the assessment of objective and subjective formal Thought and Language Disorder (TALD). *Schizophr Res*. 2014;160(1-3):216-221.
 9. Andreasen NC. Scale for the assessment of thought, language, and communication (TLC). *Schizophr Bull*. 1986;12(3):473-482.
 10. Corcoran CM, Mittal VA, Bearden CE, E Gur R, Hitczenko K, Bilgrami Z, et al. Language as a biomarker for psychosis: A natural language processing approach. *Schizophr Res*. 2020;1:S0920-9964(20)30247-4.
 11. Gupta T, Hespos SJ, Horton WS, Mittal VA. Automated analysis of written narratives reveals abnormalities in referential cohesion in youth at ultra high risk for psychosis. *Schizophr Res*. 2018;192:82-88.
 12. Yung AR, McGorry PD. The prodromal phase of first episode psychosis: past and current conceptualisations. *Schizophrenia Bulletin*. 1996; 22:353–370.
 13. Van der Gaag M, Smit F, Bechdolf A, French P, Linszen DH, Yung AR, et al. Preventing a first episode of psychosis: meta-analysis of randomized controlled prevention trials of 12 month and longer-term follow-ups. *Schizophr Res* 2013; 149: 56–62.
 14. 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.
 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. 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.
 17. 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):220–229.
 18. Goldsmith DR, Crooks CL, Walker EF, Cotes RO. An Update on Promising Biomarkers in Schizophrenia. *Focus (Am Psychiatr Publ)*. 2018;16(2):153-163.
 19. Bearden CE, Wu KN, Caplan R, Cannon TD. Thought disorder and communication deviance as predictors of outcome in youth at clinical high risk for psychosis. *J Am Acad Child Adolesc Psychiatry*. 2011;50(7):669-680.
 20. Caplan R, Guthrie D, Fish B, Tanguay PE, David-Lando G. The Kiddie Formal Thought Disorder Rating Scale: clinical assessment, reliability, and validity. *J Am Acad Child Adolesc Psychiatry*. 1989;28:408-416.
 21. Halliday MAK, Hassan R. Cohesion in Spoken and Written English. London: Longman’s; 1976.
 22. Caplan R, Guthrie D, Foy JG. Communication deficits and formal thought disorder in schizophrenic children. *J Am Acad Child Adolesc Psychiatry*. 1992;31:151-159.
 23. Bedi, G., Carrillo, F., Cecchi, G. Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophr*. 2015;1:15030.
 24. Landauer TK, Dumais ST. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol Rev* 1997;104:211–240.
 25. Natural Language Toolkit. NLTK, disponible en <http://www.nltk.org/>
 26. Bird S, Klein E, Loper E. Natural Language Processing with Python. O'Reilly Media: Sebastopol, CA, USA, 2009.

27. Demjaha A, Weinstein S, Stahl D, Day F, Valmaggia L, Rutigliano G, et al. Formal thought disorder in people at ultra-high risk of psychosis. *BJPsych Open*. 2017;3(4):165-170.
28. Murray HA. Thematic Apperception Test. Harvard University Press, 1943.
29. Liddle PF, Ngan ET, Caissie SL, Anderson CM, Bates AT, Quedsted DJ, et al. Thought and Language Index: an instrument for assessing thought and language in schizophrenia. *Br J Psychiatry*. 2002;181:326–30.
30. Gutierrez, E. D., Cecchi, G. A., Corcoran, C., & Corlett, P. Using automated metaphor identification to aid in detection and prediction of first-episode schizophrenia. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2017. p. 2923-2930.
31. Dinh ED, Gurevych I. Token-level metaphor detection using neural networks. In *Proceedings of the Fourth Workshop on Metaphor in NLP*. 2016. p. 28-33.
32. Socher R, Huval B, Manning CD, Ng AY. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. 2012. p. 1201– 1211.
33. Corcoran C. M., Carrillo F., Fernández-Slezak D., Bedi G., Klim C., Javitt D. C. et al. Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry*. 2018;17(1):67-75.
34. Rezaei N, Walker E, Wolff P. A machine learning approach to predicting psychosis using semantic density and latent content analysis. *NPJ Schizophr*. 2019;5(1):9.
35. Addington J, Liu L, Buchy L, Cadenhead, K. S., Cannon, T. D., Cornblatt, B. A, et al. North American Prodrome Longitudinal Study (NAPLS 2): The Prodromal Symptoms. *J Nerv Ment Dis*. 2015;203(5):328-335.
36. Addington J, Liu L, Brummitt K, Bearden, C. E., Cadenhead, K. S., Cornblatt, B. A. et al. North American Prodrome Longitudinal Study (NAPLS 3): Methods and baseline description. *Schizophr Res*. 2020;S0920-9964(20)30217-6.
37. Klein D, Manning CD. Accurate unlexicalized parsing. In *Proceedings of the 41st annual meeting of the association for computational linguistics*. 2003. p. 423-430.
38. Computerized Propositional Idea Density Rater. Disposable online en: www.covingtoninnovations.com/software.html
39. Cannon TD, Cadenhead K, Cornblatt B, Woods S. W., 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(1):28-37.
40. Nelson B, Yuen HP, Wood SJ, Lin A., Spi-liotacopoulos D., Bruxner A, et al. Long-term follow-up of a group at ultra high risk ("prodromal") for psychosis: the PACE 400 study. *JAMA Psychiatry*. 2013;70(8):793-802.
41. Tarbox SI, Addington J, Cadenhead KS, Cannon T. D., Cornblatt B. A., Perkins D. O, et al. Premorbid functional development and conversion to psychosis in clinical high-risk youths. *Dev Psychopathol*. 2013;25(4 Pt 1):1171-1186.
42. Addington J, Cadenhead KS, Cannon TD, Cornblatt B., McGlashan T. H., Perkins D. O. et al. North American Prodrome Longitudinal Study: a collaborative multisite approach to prodromal schizophrenia research. *Schizophr Bull*. 2007;33(3):665-672.
43. DeVylder JE, Muchomba FM, Gill KE, Ben-David S., Walder D. J., Malaspina D. et al. Symptom trajectories and psychosis onset in a clinical high-risk cohort: the relevance of subthreshold thought disorder. *Schizophr Res*. 2014;159(2-3):278-283.
44. Cornblatt BA, Carrion RE, Auther A, McLaughlin, D., Olsen R. H., John M, et al. Psychosis prevention: a modified clinical high risk perspective from the recognition and prevention (RAP) program. *Am J Psychiatry*. 2015;172:986–994.
45. Cornblatt BA. The New York high risk project to the Hillside recognition and prevention (RAP) program. *Am J Med Genet*. 2002;114(8):956-966.
46. Armando M, Pontillo M, De Crescenzo F, Mazzone L., Monducci E., Cascio N.

- L, et al. Twelve-month psychosis-predictive value of the ultra-high risk criteria in children and adolescents. *Schizophr Res.* 2015;169(1-3):186-192.
47. Rekhi G, Rapisarda A, Lee J. Impact of distress related to attenuated psychotic symptoms in individuals at ultra high risk of psychosis: Findings from the Longitudinal Youth at Risk Study. *Early Interv Psychiatry.* 2019;13(1):73-78.
48. Kotlicka-Antczak M, Karbownik MS, Stawiski K, Pawełczyk A., Żurner N., Pawełczyk T., et al. Short clinically-based prediction model to forecast transition to psychosis in individuals at clinical high risk state. *Eur Psychiatry.* 2019;58:72-79.

Correspondence to:
Ignacio Olivares Romero
Blanco Encalada Avenue 1727, Suite 302.
Santiago Downtown, Santiago.
E-mail: ignacio.olivares.ro@gmail.com
Telephone: 9 97738990