UCLA
UCLA Previously Published Works

Title

The Early Psychosis Screener (EPS): Quantitative validation against the SIPS using machine learning

Permalink

https://escholarship.org/uc/item/2q59t38j

Authors

Brodey, BB
Girgis, RR
Favorov, OV
et al.

Publication Date

2018-07-01

DOI

10.1016/j.schres.2017.11.030

Peer reviewed

The Early Psychosis Screener (EPS): Quantitative Validation against the SIPS Using Machine Learning

BB Brodey[a], **RR Girgis**[b], **OV Favorov**[c], **J Addington**[d], **DO Perkins**[e], **CE Bearden**[f], **SW Woods**[g], **EF Walker**[h], **BA Cornblatt**[i], **G Brucato**[b], **B Walsh**[j], **KA Elkin**[k], **IS Brodey**[l]

RR Girgis: Ragy.girgis@nyspi.columbia.edu; OV Favorov: favorov@email.unc.edu; J Addington: jmadding@ucalgary.ca; DO Perkins: diana_perkins@med.unc.edu; CE Bearden: cbearden@mednet.ucla.edu; SW Woods: scott.woods@yale.edu; EF Walker: psyefw@emory.edu; BA Cornblatt: cornblat@lij.edu; B Walsh: barbara.walsh@yale.edu; KA Elkin: kelkin@telesage.com; IS Brodey: brodey@email.unc.edu

[a]TeleSage, Inc., 201 East Rosemary St., Chapel Hill, NC 27514, USA

[b]New York State Psychiatric Institute, 1051 Riverside Drive, Unit 31, New York, NY 10032, USA

[c]Department of Biomedical Engineering, University of North Carolina at Chapel Hill, 152 MacNider Hall, Campus Box 7575, Chapel Hill, NC 27599, USA

[d]Hotchkiss Brain Institute, Department of Psychiatry, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6 Canada

[e]Department of Psychiatry, School of Medicine, University of North Carolina at Chapel Hill, 101 Manning Dr, Chapel Hill, NC 27514, USA

[f]Departments of Psychiatry and Biobehavioral Sciences and Psychology, University of California Los Angeles, 757 Westwood Plaza, Los Angeles, CA 90095, USA

[g]PRIME Psychosis Prodrome Research Clinic, Connecticut Mental Health Center B-38, 34 Park Street, New Haven, CT 06519, USA

[h]Departments of Psychology and Psychiatry, Emory University, 36 Eagle Row, Atlanta, GA 30322, USA

Corresponding Author: bb@telesage.com, Phone: 866-942-8849, Fax: 919-942-0036.

[i]Department of Psychiatry Research, The Zucker Hillside Hospital, 75-59 263rd St., Glen Oaks, New York 11004, USA

[j]PRIME Psychosis Prodrome Research Clinic, Connecticut Mental Health Center B-38, 34 Park Street, New Haven, CT 06519, USA

[k]TeleSage, Inc., 201 East Rosemary St., Chapel Hill, NC 27514, USA

[l]TeleSage, Inc., 201 East Rosemary St., Chapel Hill, NC 27514, USA

Abstract

Machine learning techniques were used to identify highly informative early psychosis self-report items and to validate an early psychosis screener (EPS) against the Structured Interview for Psychosis-risk Syndromes (SIPS). The Prodromal Questionnaire– Brief Version (PQ-B) and 148 additional items were administered to 229 individuals being screened with the SIPS at 7 North American Prodrome Longitudinal Study sites and at Columbia University. Fifty individuals were found to have SIPS scores of 0, 1, or 2, making them clinically low risk (CLR) controls; 144 were classified as clinically high risk (CHR) (SIPS 3–5) and 35 were found to have first episode psychosis (FEP) (SIPS 6). Spectral clustering analysis, performed on 124 of the items, yielded two cohesive item groups, the first mostly related to psychosis and mania, the second mostly related to depression, anxiety, and social and general work/school functioning. Items within each group were sorted according to their usefulness in distinguishing between CLR and CHR individuals using the Minimum Redundancy Maximum Relevance procedure. A receiver operating characteristic area under the curve (AUC) analysis indicated that maximal differentiation of CLR and CHR participants was achieved with a 26-item solution (AUC = 0.899±0.001). The EPS-26 outperformed the PQ-B (AUC = 0.834±0.001). For screening purposes, the self-report EPS-26 appeared to differentiate individuals who are either CLR or CHR approximately as well as the clinician-administered SIPS. The EPS-26 may prove useful as a self-report screener and may lead to a decrease in the duration of untreated psychosis. A validation of the EPS-26 against actual conversion is underway.

Keywords

SIPS; PQ-B; NAPLS; psychosis; prodromal; schizophrenia; screener; machine learning

1. Introduction

Clinicians that attempt to ameliorate the symptoms of schizophrenia and other psychoses, after the symptoms have developed, have been met with limited success. A newer approach is identifying individuals who are at increased risk of developing psychotic disorders in order to prevent progression of the illness and to decrease the duration of untreated psychosis (Kline and Schiffman, 2014). The Structured Interview for Psychosis-risk Syndromes (SIPS) was developed to identify clinically high risk (CHR) individuals in order to evaluate the natural history of the illness during the prodromal period and to identify interventions that could help prevent progression (Miller et al., 1999, 2002; McGlashan et al., 2001). The SIPS is the "gold standard" early psychosis assessment in North America, but

it is also a structured interview that takes about 90 minutes to administer and requires extensive training to assure high inter-rater reliability (Miller et al., 2003). For these reasons, its use is often restricted to research centers. The Prodromal Questionnaire – Brief Version (PQ-B) was developed a few years later in order to simplify the process of identifying individuals who are CHR (Loewy et al., 2005, 2011a). Although other instruments have been developed for screening purposes, the PQ-B is the most researched self-report screener (Jarrett et al., 2012; Kline et al., 2012a, 2012b; Loewy et al., 2011b; Okewole et al., 2015). Despite the research behind it, the high false positive rate of the PQ-B may make it unsuitable for widespread use as a screener in many populations (Kline et al., 2012b; Xu et al., 2016). Given the low prevalence of early psychosis in the general population, it is desirable to have a more specific screener for early psychosis to promote early intervention (Cohen and Marino, 2013; Comparelli et al., 2014).

In an earlier project, TeleSage developed a self-report item bank to serve as the foundation for developing an early psychosis screener (EPS). We assembled a panel of experts and implemented a rigorous survey item development, modification, and selection process. This process included 40 participants and up to five rounds of cognitive interviewing per item (Willis, 2005). We identified a subset of 148 items that were well understood by prodromal individuals and that our expert panel believed would cover the breadth of concepts associated with the prodromal period and early psychosis. After removing items from the survey that were unnecessary for our analyses (see section 3.1.1.), we were left with 124 items for the machine learning analysis.

In initiating the present study, we wanted to validate an EPS instrument based on the rigor of the established North American Prodrome Longitudinal Study (NAPLS) clinics and the Center of Prevention and Evaluation (COPE) clinic at Columbia University. We used machine learning techniques and the response sets gathered from established prodromal sites to maximize our ability to develop a useful EPS.

Our hypothesis is that machine learning techniques can be used to select a minimal subset of the 124 self-report items that can be used to identify with high sensitivity and specificity individuals who are at clinically high risk for developing psychosis.

2. Methods

2.1. Participants

TeleSage, Inc. partnered with the Columbia University COPE Clinic and seven NAPLS research sites, located at Emory University, University of Calgary, UCLA, UCSD, UNC-Chapel Hill, Yale University, and Zucker Hillside Hospital. All of the clinical participants in this study were recruited from these eight sites. Overall, we recruited 229 participants (demographic information is presented in Table 1). The recruitment procedures for the NAPLS sites and COPE have been comprehensively described in the literature (Addington, 2012; Brucato 2017).

IRB approval was obtained for all sites at their host institutions, and all participants provided IRB-approved informed consent. At the NAPLS sites and at the COPE clinic the CLR, CHR,

and FEP groups were defined by the Criteria of Psychosis-risk Syndromes (COPS), contained in the SIPS (McGlashan et. al 2010). Exclusion criteria included attenuated positive symptoms better accounted for by another psychiatric condition, past or present full-blown psychosis, I.Q. < 70, medical conditions known to affect the central nervous system, and current serious risk of harm to self or others. Eligible participants in this study were recruited from a pool of patients who were already receiving a SIPS evaluation for a primary CHR-related study (see Miller, 2003 for a description of the SIPS assessment procedures). Individuals who received the SIPS were asked to participate in the EPS study. Participants who scored a 0, 1, or 2 on the all of the SIPS positive symptoms were placed in the clinically low risk (CLR) group. Participants who scored a 3, 4, or 5 on one or more of the SIPS positive symptoms were placed in the CHR group. Participants scoring 6 on any of the SIPS positive symptoms were placed in the active psychosis (FEP) group. All participants completed paper assessments including 9 demographics items, our 148 test items, and the PQ-B.

2.2. Analytical Procedures

The analyses were performed on the participants' answers to the questionnaire items. The goal of this study was to develop the most effective computational procedure for reducing the Likert scale survey answers of a tested individual to a single quantitative metric, or a score, that could be used to infer that individual's SIPS class identity. The simplest such metric is a linear sum of answers to all the items:

$$M_{LS} = \sum_{i \in Q} L_i \quad (1)$$

where Q is a set of questionnaire items and L_i is the Likert scale answer to the ith item.

The linear sum metric M_{LS} is limited in its representational power, however, since it treats all the items as contributing uniformly to SIPS class estimation. In the supplementary information published online, we consider more versatile linear and nonlinear metrics but find that their CLR vs. CHR discriminatory performance is not superior to the performance of the linear sum metric M_{LS}. Consequently, we chose M_{LS} as the best metric suited for our screener.

The capacity of M_{LS} to accurately predict which SIPS class a tested individual belongs to based on his/her EPS questionnaire answers was evaluated using receiver operating characteristic (ROC) analyses. The classification accuracy was expressed as the area under the ROC curve (AUC). AUC values can range between 0.5 (for classifiers whose performance is completely random) and 1 (for perfectly accurate classifiers).

Two analytical approaches were used to identify those among the original list of 124 survey items that could be safely omitted from the final list. The first approach was spectral clustering, which was used to identify clusters of the questionnaire items with distinctly different patterns of answers among individuals belonging to CLR, CHR, and FEP groups (Shi and Malik, 2000; Ng et al., 2001; von Luxburg, 2007). We measured the similarity between different items by computing their correlation coefficient over all four groups of

subjects. Such pairwise correlation coefficients make up a similarity matrix S. Importantly, no information about the subjects' group membership was used in computing the correlation coefficients and, therefore, in creating the similarity matrix S. This similarity matrix S is used to construct normalized graph Laplacian matrix:

$$L_{NCut} = D^{-1/2} \cdot (D - S)^{-1/2} \quad (2)$$

where D is a diagonal matrix, in which $D_{ii} = \Sigma_j S_{ij}$. To determine how many distinct groups are present among the items, we compute and plot "eigengaps" between consecutive eigenvalues $\lambda_1 \ldots \lambda_N$ of L_{NCut} matrix (the ith eigengap is defined as a difference $\lambda_i = \lambda_{i+1} - \lambda_i$; with the first eigengap, λ_1, set to zero). In general, if a dataset has K distinct clusters, the eigengap plot will have an outstanding eigengap in the K position (λ_K) and also likely to the left of it, but not to the right. The corresponding K^{th} eigenvector sorts all the items into two groups, which can be seen by plotting that eigenvector. (For an in-depth description of the spectral clustering approach and procedures, see Supplementary Information.) It should be pointed out that our spectral clustering approach to partitioning the 124 items into smaller subsets does not rely at all on the membership of subjects in the CLR, CHR, or FEP groups. The purpose of this partitioning was not to select the more discriminative items, but to improve the items-to-participants ratio, so as to increase our power to identify the most informative items in each reduced subset.

The second analytical approach was Minimum Redundancy Maximum Relevance (mRMR) analysis. mRMR is an effective feature selection approach used in machine learning, which addresses the well-known problem that combinations of individually good variables do not necessarily lead to good classification performance by aiming to maximize the joint class dependency of the selected variables by minimizing the redundancies among the selected most relevant variables (Peng et al. 2005). We used the mRMR procedure to sort N given questionnaire items by incrementally selecting the maximally relevant items while avoiding the redundant ones. Accordingly, the mth item x_m chosen for inclusion in the set of already selected items, S, must satisfy the following condition:

$$\max_{x_j \in X - S_{m-1}} \left[I(x_j, c) - \frac{1}{m-1} \sum_{x_i \in S_{m-1}} I(x_j; x_i) \right], \quad (3)$$

where X is the entire set of N items; c is the SIPS class variable; x_i is the ith selected item; and I is mutual information. In other words, the item that has the maximum difference between its mutual information with the class variable and the average mutual information with the items in S will be chosen next.

3. Results

3.1. Item Selection

3.1.1. General Considerations—To avoid potentially spurious differentiations based on age, gender, race, education, employment, and friendships, we removed the items on demographic information. Next, although we gathered detailed data on participants' alcohol and drug use, drug usage varied greatly and no particular drug other than marijuana was regularly endorsed. Additionally, we were aware of the potential inaccuracy of self-report drug use data. To avoid potential complications, which we could not address due to the limited number of participants with drug use, we removed items on alcohol and drug use prior to the analysis. Finally, we removed 12 items that were not applicable to all participants (i.e., specific work or study related items). In all, we were left with 124 items.

3.1.2. Spectral Clustering—Making mRMR sort many more items ($n = 124$) than the number of CLR subjects ($n = 50$) would reduce that algorithm's effectiveness. To avoid such an item/subject imbalance, we first used Spectral Clustering to split the 124 items into smaller-size groups of similarly behaving items and then used mRMR separately on each of those groups. To determine whether any of the 124 items formed distinct groups with regard to their coincident variations among the studied individuals, we computed eigenvalues of the normalized graph Laplacian matrix L_{NCut} (equation 2) and plotted their eigengaps (Figure 1, graph A). This eigengap plot revealed just one outstanding eigengap: λ_2. Following the rule that the rightmost outstanding eigengap indicates the number of distinct clusters, we conclude that the 124 items formed two distinct clusters with regard to how participants answered them.

To find out how the 124 items were divided into the 2 clusters, we plotted the 2nd eigenvector, which performs this division in Figure 1 (graph B). In this plot, the height of each bar indicates how well each item fits into either of the two groups, while the positive/negative sign of each bar indicates to which group each item was assigned. Significantly, an overwhelming majority of the 61 positive symptom items (Group P) target either psychosis or mania. In contrast, the 63 negative symptom items (Group N) predominantly target depression, anxiety, and social and general work/school functioning. (Figure S1 in Supplemental Information shows that the membership of individual items in the two groups is highly reproducible.)

3.1.3. mRMR—The mRMR scoring was performed separately on the 61 Group P items and the 63 Group N items with 50 CLR and 144 CHR subjects. Using the bootstrapping with replacement approach, computation of the mRMR scores of each group of items was repeated 200 times, and the items were sorted according to their average scores. These average mRMR scores are plotted in Figure 2 (graphs A and B).

ROC curves were constructed for progressively more inclusive subsets of items with the highest average mRMR scores to determine the usefulness of various items in a group for distinguishing between CLR and CHR individuals. This was done separately for each group. Figure 2 (graph C) plots AUC of these ROC curves as a function of the number of items used to construct the curves. The plot shows that for Group P, after the top 26 items were

selected by mRMR, adding more items did not improve the classification performance, but added noise and decreased the AUC of the item pool. For the top 26 items, AUC = 0.899 ± 0.001. For Group N, AUC reached its peak of 0.846 ± 0.001 at 6 items and declined progressively with further addition of more items. Thus we reduced the candidate set of items for the screener from 124 to 32 (i.e., 26 from Group P and 6 from Group N).

For the second round of item selection, we repeated the mRMR procedure on the combined set of the chosen 32 items but found that the peak AUC of 0.900 ± 0.001 was reached only when using all 32 items (Figure 2, graph D). Since we obtained the same AUC with just 26 items from Group P, we conclude that the EPS can use just these 26 items (the EPS-26). (A full copy of the EPS-26 can be found in the on-line addendum associated with this manuscript. Figure S2 in Supplemental Information addresses the question of how definitive the selection is of the final 26 items. It shows that the entire pool of discriminatively useful items is around 30, but only 20 of those items are most useful, whereas the remaining ones make only minor contributions.)

3.2. EPS-26 Discriminative Performance

In addition to CLR and CHR individuals, we tested EPS-26 on participants suffering from psychosis (the FEP group). Table 2 lists average ROC AUC obtained by pairing all of the 3 groups against each other (using bootstrapping with replacement 1000 times for each pair). According to this table, EPS-26 discriminates comparably well between CLR and CHR, and CLR and FEP, but shows little discrimination between CHR and FEP.

We compared the discriminative performance of EPS-26 on our CLR and CHR sample with that of another commonly used screener, the PQ-B (Loewy et al., 2011a). The PQ-B items were scored in their T/F format. Figure 3 plots superimposed ROC curves for PQ-B and EPS-26, revealing that EPS-26 performance is superior to PQ-B, whose AUC = 0.834 ± 0.001. The difference between AUC of EPS-26 and PQ-B is statistically significant ($p = 0.0069$; determined using the statistical comparison method of Hanley and McNeil, 1983).

4. Discussion

This paper uses machine learning techniques to establish a 26-item early psychosis screener (the EPS-26) from a rigorously developed and comprehensive item bank. During the development of the EPS-26, we eliminated items that had the potential to sort individuals based on criteria that were unrelated to the desired trait. We sorted the remaining items into two groups that appeared to represent different factors, and we ranked the items based on how informative they were in sorting the groups. We selected the items that were most informative and eliminated items that added noise without improving the ability of the group of items to differentiate between the two groups. Throughout this process, we employed established techniques to avoid overfitting the data.

Our hope is that the EPS-26 will be used to identify individuals who should be referred to specialty providers for further in-person evaluation for prodromal status. Based on the EPS-26 ROC curve presented in Figure 3 and a hypothetical incidence of CHR status at an outpatient behavioral health clinic, we can imagine several scenarios (summarized in Table

3). Although sensitivity exactly equals selectivity at 83 in this study, we used a sensitivity of 80 in the scenarios below.

Scenario 1: A clinic screens new clients who seem to have unusual thoughts or perceptions, or who exhibit social withdrawal. Because the clinic only screens these clients, and not everyone who is a new client, we assume that 20% of this population is CHR. We also assume that the clinicians in this clinic want a self-report screener that can identify 80% of the people who qualify as CHR (80% sensitivity). Based on this scenario and the actual ROC curve for the EPS-26 (Figure 3), for every true CHR client identified, 0.9 clients would be falsely identified as CHR (false positive) and 0.16 would be falsely identified as CLR (false negative). In our view, selecting a screener with high sensitivity in a population with high incidence might be clinically useful.

Scenario 2: While still retaining a desired sensitivity of 80%, this scenario is different from the first in that every new client is screened using the self-report screener. Thus, we will assume that only 5% of these clients are actually CHR. Now based on the ROC curve for the EPS-26, for every true positive client identified, 3.75 "false positives" will also need to be evaluated. This scenario might result in excessive clinical burden; thus, selecting a high sensitivity in a population with low incidence may not be clinically useful.

Scenario 3: This final scenario retains the population characteristics of Scenario 2 (5% CHR), but decreases the sensitivity of the assessment to 50%. Based on the ROC curve for the EPS-26, for every true positive client identified, 1.2 CLR "false positive" clients will also be considered for further evaluation, and 1 CHR client will be wrongly identified as CLR (false negative). Considering this scenario, selecting a screener with lower sensitivity in a population with low incidence might be clinically useful.

The scenarios above reflect hypothetical populations in which each respondent answers as accurately and as truthfully as possible. However, a failing of self-report assessments is that they are, in general, prone to purposeful manipulation. Some individuals with help-seeking behaviors may attempt to fake symptoms. Other individuals wishing to demonstrate that they are well (e.g., to enter the military) might attempt to minimize symptoms. Fortunately, our response set uses a Likert scale, and it is already clear that there are orderly relationships between certain responses. Provided that the outcome variables are known, our expectation is that as the EPS-26 is used more widely, it will be possible to identify and report patterns that invalidate the assessment.

Along with the Likert scale response set, two additional benefits of the EPS-26 are noteworthy. First, to the greatest extent possible, we designed the individual EPS items so that each one asks about a single granular concept. If we look at the individual item endorsement patterns, it should be possible to determine which granular concepts and clusters are associated with CHR status. An added benefit of this work is that, based on the results presented in Table 2, the ability of the EPS-26 to identify CHR status appears to be equivalent to the ability of the EPS-26 to identify early psychosis. The EPS may thus provide a useful tool for shortening the duration of untreated psychosis.

We were able to create a self-report assessment that accurately predicts SIPS CLR and CHR categories, but this study has several important limitations. We only evaluated people who were referred to a specialty early psychosis research center for evaluation and who chose to receive the evaluation. Exclusionary criteria included attenuated positive symptoms better accounted for by another psychiatric condition, past or present full-blown psychosis, I.Q. < 70, a medical condition known to affect the central nervous system, and current serious risk of harm to self or others. Despite these exclusionary criteria we were able to include participants who had more minor general psychopathology in the CLR CHR populations, but we can only report on the population being evaluated at the NAPLS and COPE sites. In the future, we hope to be able to report on the use of the EPS-26 in broader populations. In addition, we remain concerned that although the gold standard SIPS has good sensitivity (about 95%), only 19.6% of CHR individuals actually convert (Webb et al., 2015). This is a limitation in the design of this study, since it is not possible for any assessment to be superior to the gold standard assessment that is being used for its validation. For this reason, future work with the EPS-26 will include validation against true conversion rates.

5. Conclusions

The machine learning techniques we applied in this study enabled us to successfully select 26 self-report items that identify individuals who are at clinically high risk for psychosis with high sensitivity and specificity. Overall, the sensitivities and specificities that we achieved using the EPS-26 were superior to those obtained using the PQ-B in the same sample. Our hope is that the EPS-26 will be used for widespread screening in clinical settings, as a self-report alternative to the SIPS. Extensive screening with a highly specific self-report screener, such as the EPS-26, might lead to the early identification of at-risk individuals and spur research on effective interventions. Validation of the EPS-26 against true conversion rates will be the goal of future work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Abbreviations

AUC	area under the curve
CHR	clinically high risk
CLR	clinically low risk
COPE	Center of Prevention and Evaluation

EPS	early psychosis screener
FEP	first episode psychosis
mRMR	Minimum Redundancy Maximum Relevance
NAPLS	North American Prodrome Longitudinal Study
PQ-B	Prodromal Questionnaire – Brief Version
ROC	receiver operating characteristic
SIPS	Structured Interview for Psychosis-risk Syndromes

References

Addington J, Cadenhead KS, Cornblatt BA, Mathalon DH, McGlashan TH, Perkins DO, Seidman LJ, Tsuang MT, Walker EF, Woods SW, Addington JA, Cannon TD. 2012; North American Prodrome Longitudinal Study (NAPLS 2): Overview and recruitment. Schizophr Res. 142:77–82. DOI: 10.1016/j.schres.2012.09.012 [PubMed: 23043872]

Brucato1*†, G, Masucci1†, MD, Arndt1, LY, Ben-David1, S, Colibazzi1, T, Corcoran1, CM, Crumbley2, AH, Crump1, FM, Gill3, KE, Kimhy1, D, Lister1, A, Schobel4, SA, Yang5, LH, Lieberman1, JA, Girgis1, RR. Psychological Medicine. Baseline demographics, clinical features and predictors of conversion among 200 individuals in a longitudinal prospective psychosis-risk cohort; 1–13. © Cambridge University Press 2017

Cohen CI, Marino L. 2013; Racial and ethnic differences in the prevalence of psychotic symptoms in the general population. Psychiatr Serv. 64(11):1103–1109. DOI: 10.1176/appi.ps.201200348 [PubMed: 23904054]

Comparelli A, De Carolis A, Emili E, Rigucci S, Falcone I, Corigliano V, Curto M, Trovini G, Dehning J, Kotzalidis GD, Girardi P. 2014; Basic symptoms and psychotic symptoms: their relationships in the at risk mental states, first episode and multi-episode schizophrenia. Compr Psychiatry. 55:785–791. DOI: 10.1016/j.comppsych.2014.01.006 [PubMed: 24556516]

Hanley JA, McNeil BJ. 1983; A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 148:839–843. DOI: 10.1148/radiology. 148.3.6878708 [PubMed: 6878708]

Jarrett M, Craig T, Parrott J, Forrester A, Winton-Brown T, Maguire H, McGuire P, Valmaggia L. 2012; Identifying men at ultra high risk of psychosis in a prison population. Schizophr Res. 136(1–3):1–6. DOI: 10.1016/j.schres.2012.01.025 [PubMed: 22330178]

Kline E, Wilson C, Ereshefsky S, Tsuki T, Schiffman J, Pitts S, Reeves G. 2012a; Convergent and discriminant validity of attenuated psychosis screening tools. Schizophr Res. 134(1):49–53. DOI: 10.1016/j.schres.2011.10.001 [PubMed: 22036199]

Kline E, Wilson C, Ereshefsky S, Denenny D, Thompson E, Pitts SC, Bussell K, Reeves G, Schiffman J. 2012b; Psychosis risk screening in youth: A validation study of three self-report measures of attenuated psychosis symptoms. Schizophr Res. 141(1):72–77. DOI: 10.1016/j.schres.2012.07.022 [PubMed: 22921375]

Kline E, Schiffman J. 2014; Psychosis risk screening: a systematic review. Schizophr Res. 158:11–18. DOI: 10.1016/j.schres.2014.06.036 [PubMed: 25034762]

Loewy RL, Bearden CE, Johnson JK, Raine A, Cannon TD. 2005; The prodromal questionnaire (PQ): Preliminary validation of a self-report screening measure for prodromal and psychotic syndromes. Schizophr Res. 79:117–125. DOI: 10.1016/j.schres.2005.03.007 [PubMed: 16276559]

Loewy RL, Pearson R, Vinogradov S, Bearden CE, Cannon TD. 2011a; Psychosis risk screening with the Prodromal Questionnaire – Brief Version (PQ-B). Schizophr Res. 129:42–49. DOI: 10.1016/j.schres.2011.03.029 [PubMed: 21511440]

Loewy RL, Therman S, Manninen M, Huttunen MO, Cannon TD. 2011b; Prodromal psychosis screening in adolescent psychiatry clinics. Early Interv Psychiatry. 6:69–75. DOI: 10.1111/j. 1751-7893.2011.00286.x [PubMed: 21883972]

McGlashan, TH, Miller, TJ, Woods, SW, Hoffman, RE, Davidson, L. A scale for the assessment of prodromal symptoms and states. In: Miller, TJ, Mednick, SA, McGlashan, TT, Liberger, J, Johannessen, JO, editors. Early Intervention in Psychotic Disorders. Dordrecht, The Netherlands: Kluwer Academic Publishers; 2001. 135–149.

McGlashan, et al. 2010

Miller TJ, McGlashan TH, Woods SW, Stein K, Driesen N, Corcoran CM, Hoffman R, Davidson L. 1999; Symptom assessment in schizophrenic prodromal states. Psychiatr Q. 70(4):273–287. DOI: 10.1023/A:1022034115078 [PubMed: 10587984]

Miller TJ, McGlashan TH, Rosen JL, Somjee L, Markovich PJ, Stein K, Woods SW. 2002; Prospective diagnosis of the initial prodrome for schizophrenia based on the Structured Interview for Prodromal Syndromes: Preliminary evidence of interrater reliability and predictive validity. Am J Psychiatry. 159:863–865. DOI: 10.1176/appi.ajp.159.5.863 [PubMed: 11986145]

Miller TJ, McGlashan TH, Rosen JL, Cadenhead K, Ventura J, McFarlane W, Perkin DO, Pearlson GD, Woods SW. 2003; 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. 29(4):703–715. DOI: 10.1093/oxfordjournals.schbul.a007040 [PubMed: 14989408]

Ng, AY; Jordan, MI; Weiss, Y. On spectral clustering: Analysis and an algorithm. In: Dietterich, TG; Becker, S; Ghahramani, Z, editors. NIPS'01 Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic; Paper presented at NIPS; 2001; Vancouver, British Columbia. Cambridge, MA: MIT Press; 2001. 849–856.

Okewole AO, Ajogbon D, Adenjii AA, Omotoso OO, Awhangansi SS, Fasokun ME, Agbola AA, Oyekanmi AK. 2015; Psychosis risk screening among secondary school students in Abeokuta, Nigeria: Validity of the Prodromal Questionnaire – Brief Version (PQ-B). Schizophr Res. 164(1–3):281–282. DOI: 10.1016/j.schres.2015.01.006 [PubMed: 25640525]

Peng H, Long F, Ding C. 2005; Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell. 27(8):1226–1238. DOI: 10.1109/TPAMI.2005.159 [PubMed: 16119262]

Shi J, Malik J. 2000; Normalized cuts and image segmentation. IEEE Trans Pattern Anal Mach Intell. 22(8):888–905. DOI: 10.1109/34.868688

von Luxburg U. 2007; A tutorial on spectral clustering. Stat Comput. 17(4):395–416. DOI: 10.1007/s11222-007-9033-z

Webb JR, Addington J, Perkins DO, Bearden CE, Cadenhead KS, Cannon TD, Barbara AC, Heinssen RK, Seidman LJ, Tarbox SI, Tsuang MT, Walker EF, McGlashan TH, Woods SW. 2015; Specificity of incident diagnostic outcomes in patients at clinical high risk for psychosis. Schizophr Bull. 41(5):1066–1075. DOI: 10.1093/schbul/sbv091 [PubMed: 26272875]

Willis, G. Cognitive Interviewing: A tool for Improving Questionnaire Design. Thousand Oaks CA: Sage Publications; 2005.

Xu, LH; Zhang, TH; Zheng, LN; Tang, YY; Luo, XG; Sheng, JH; Wang, JJ. Psychometric properties of Prodromal Questionnaire – Brief Version among Chinese help-seeking individuals. PLoS One. 2016.
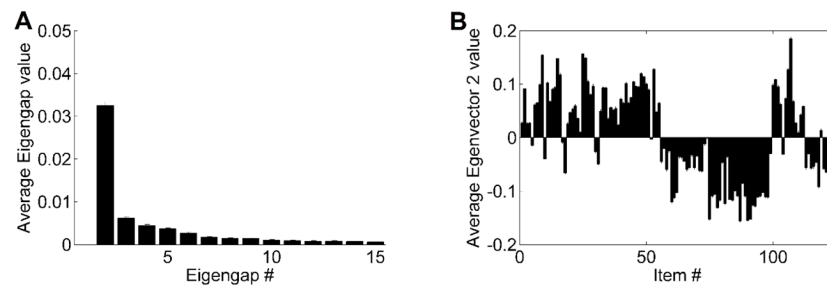
Figure 1. Spectral Clustering analysis of questionnaire items
(A) Eigengap plot of the differences in magnitude between successive eigenvalues of the normalized graph Laplacian matrix, L_{NCut}, of the similarity matrix, S, constructed for the 124 items (equation 2). This plot is an average of 100 eigengap plots, each of which was generated on a different randomly selected subsample of the study participants. Each such subsample comprised 50 subjects per group, drawn at random (with replacement) from among all subjects in each group. There is just one outstanding eigengap in this plot, λ_2, between eigenvalues 2 and 3, indicating that the items form two prominent clusters. **(B)** Average 2nd eigenvector plot, showing average of the 2nd eigenvectors computed for the same 100 random subsamples of the study participants. The plot shows the graded membership of the 124 items in the two clusters indicated by the eigengap plot.
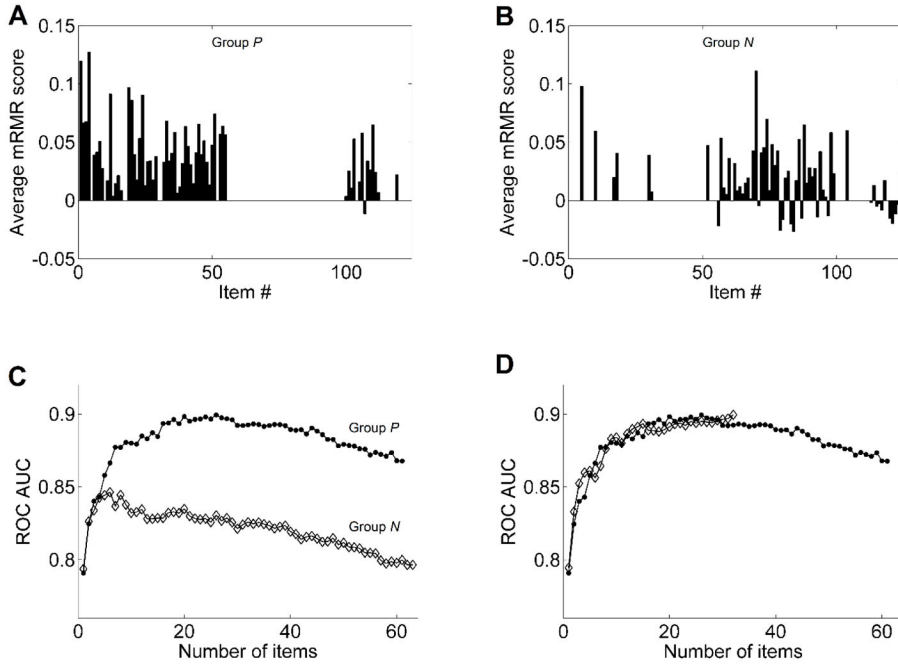
Figure 2. MRMR analysis of questionnaire items

(A) The average mRMR scores computed for the 61 items in the *P* group. **(B)** The average mRMR scores computed for the 63 items in the *N* group. **(C)** Average ROC AUC plotted as a function of the number of items with the highest average mRMR scores taken either from Group *P* (filled circles) or Group *N* (open diamonds). Each plotted AUC is a bootstrapping average of 1000 ROC curves, each of which was generated from a different set of 194 subjects drawn at random (with replacement) from both CLR and CHR groups. **(D)** Average ROC AUC plotted as a function of the number of items with the highest average mRMR scores taken from among the top 26 Group *P* items and 6 Group *N* items (open diamonds). Each plotted AUC is a bootstrapping average of 1000 ROC curves. For a comparison, this AUC curve is plotted superimposed over the AUC curve of the 26 Group *P* items (closed circles), reproduced from panel C.
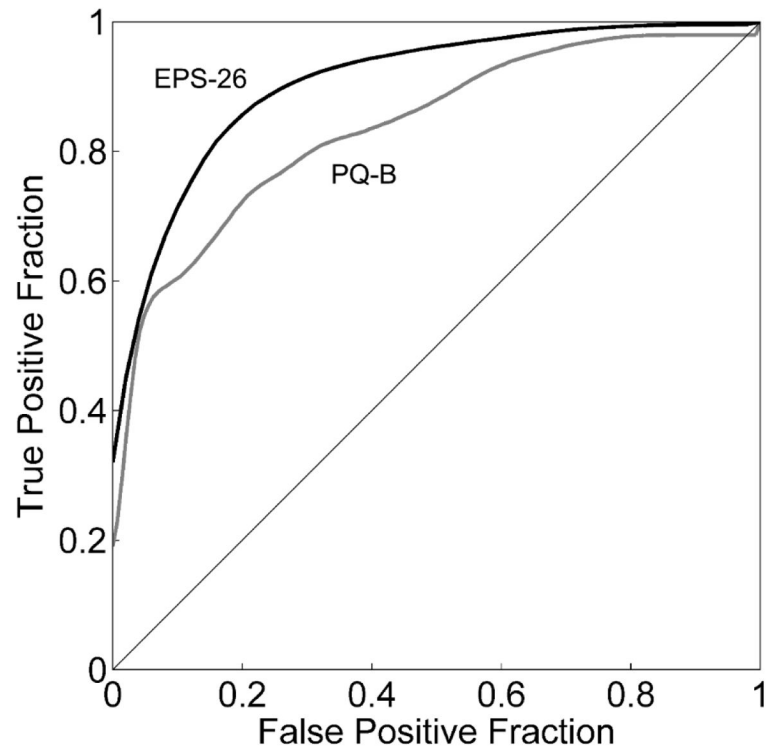
Figure 3. ROC curves for the PQ-B and EPS-26
ROC curves for discriminating between CLR subjects and CHR subjects using PQ-B (gray curve) and EPS-26 (black curve) classifiers.

Table 1

Demographics of the studied groups.

Group	n	Age (years)	Female[a]	White[a]	Black[a]	Asian[a]	Hispanic[a]	Other[a]
CLR	50	20.1±4.0	26.0	48.2	16.1	7.1	12.5	16.1
CHR	144	20.7±4.8	42.4	53.2	16.7	9.0	9.6	11.5
FEP	35	22.6±4.6	45.7	54.1	21.6	5.4	8.1	10.8

[a]Data reported as percentages of the assigned group.

Table 2

Average ROC AUC obtained by pairing all 3 groups against each other

Group	Clinical Low Risk	Clinical High Risk	Active Psychosis
Clinical Low Risk	-	0.899±0.001	0.898±0.001
Clinical High Risk	0.899±0.001	-	0.614±0.002
Active Psychosis	0.898±0.001	0.614±0.002	-

Table 3

Confusion matrices for 3 scenarios involving different choices of EPS-26 classification threshold and/or prevalence of CHR in the population.

Scenario 1: 80% sensitivity; SIPS = CHR in 20% of population.		
	SIPS = CHR	**SIPS = CLR**
EPS-26 = CHR	16%	14%
EPS-26 = CLR	4%	66%

Scenario 2: 80% sensitivity; SIPS = CHR in 5% of population.		
	SIPS = CHR	**SIPS = CLR**
EPS-26 = CHR	4%	15%
EPS-26 = CLR	1%	80%

Scenario 3: 50% sensitivity; SIPS = CHR in 5% of population.		
	SIPS = CHR	**SIPS = CLR**
EPS-26 = CHR	2.5%	3%
EPS-26 = CLR	2.5%	92%