Schizophrenia Research 208 (2019) 390–396

Contents lists available at ScienceDirect

Schizophrenia Research

journal homepage: www.elsevier.com/locate/schres

The Early Psychosis Screener for Internet (EPSI)-SR: Predicting 12 month psychotic conversion using machine learning

B.B. Brodey [a,*], R.R. Girgis [b], O.V. Favorov [c], C.E. Bearden [d], S.W. Woods [e], J. Addington [f], D.O. Perkins [g], E.F. Walker [h], B.A. Cornblatt [i], G. Brucato [b], S.E. Purcell [a], I.S. Brodey [j], K.S. Cadenhead [k]

[a] TeleSage, Inc., 201 East Rosemary St., Chapel Hill, NC 27514, USA
[b] New York State Psychiatric Institute, 1051 Riverside Drive, 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] Departments of Psychiatry and Biobehavioral Sciences and Psychology, University of California Los Angeles, 757 Westwood Plaza, Los Angeles, CA 90095, USA
[e] PRIME Psychosis Prodrome Research Clinic, Connecticut Mental Health Center B-38, 34 Park Street, New Haven, CT 06519, USA
[f] Hotchkiss Brain Institute, Department of Psychiatry, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6, Canada
[g] Department of Psychiatry, School of Medicine, University of North Carolina at Chapel Hill, 101 Manning Dr, Chapel Hill, NC 27514, USA
[h] Departments of Psychology and Psychiatry, Emory University, 36 Eagle Row, Atlanta, GA 30322, USA
[i] Department of Psychiatry Research, The Zucker Hillside Hospital, 75-59 263rd St., Glen Oaks, New York 11004, USA
[j] Department of English and Comparative Literature, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA
[k] Department of Psychiatry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0810, USA

ARTICLE INFO

Article history:
Received 4 April 2018
Received in revised form 7 November 2018
Accepted 16 January 2019
Available online 15 February 2019

Keywords:
SIPS
PQ-B
Schizophrenia
First episode psychosis
Attenuated psychotic syndrome
Prodromal psychosis

ABSTRACT

Introduction: A faster and more accurate self-report screener for early psychosis is needed to promote early identification and intervention.
Methods: Self-report Likert-scale survey items were administered to individuals being screened with the Structured Interview for Psychosis-risk Syndromes (SIPS) and followed at eight early psychosis clinics. An a priori analytic plan included Spectral Clustering Analysis to reduce the item pool, followed by development of Support Vector Machine (SVM) classifiers.
Results: The cross-validated positive predictive value (PPV) of the EPSI at the default cut-off (76.5%) exceeded that of the clinician-administered SIPS (68.5%) at separating individuals who *would not* convert to psychosis within 12 months from those who either *would* convert within 12 months or who *had already* experienced a first episode psychosis (FEP). When used in tandem with the SIPS on clinical high risk participants, the EPSI increased the combined PPV to 86.6%. The SVM classified as FEP/converters only 1% of individuals in non-clinical and 4% of clinical low risk populations. Sensitivity of the EPSI, however, was 51% at the default cut-off.
Discussion: The EPSI identifies, comparably to the SIPS but in less time and with fewer resources, individuals who are either at very high risk to develop a psychotic disorder within 12 months or who are already psychotic. At its default cut-off, EPSI misses 49% of current or future psychotic cases. The cut-off can, however, be adjusted based on purpose. The EPSI is the first validated assessment to predict 12-month psychotic conversion. An online screening system, www.eps.telesage.org, is under development.

1. Introduction

Waiting to treat symptoms of schizophrenia and other psychoses until well after symptoms have developed limits the possibility of successful treatment. A newer approach is to identify individuals who are at increased risk of developing psychotic disorders in order to prevent progression to frank illness and reduce the associated functional disability (Kline and Schiffman, 2014). The Structured Interview for Psychosis-Risk Syndromes (SIPS) was developed to identify individuals at clinical high risk (CHR) for psychosis, to evaluate the natural history of the illness during the prodromal period, and to assess response to interventions to prevent progression (Miller et al., 1999, 2002; McGlashan et al., 2001). The SIPS semi-structured interview is the "gold standard" early psychosis assessment in North America, yet it takes about 90 min to administer. In addition, extensive training and certification are required to assure high inter-rater reliability (Miller et al., 2003).

* Corresponding author.
E-mail addresses: bb@telesage.com (B.B. Brodey), Ragy.Girgis@nyspi.columbia.edu (R.R. Girgis), favorov@email.unc.edu (O.V. Favorov), cbearden@mednet.ucla.edu (C.E. Bearden), scott.woods@yale.edu (S.W. Woods), jmadding@ucalgary.ca (J. Addington), diana_perkins@med.unc.edu (D.O. Perkins), psyefw@emory.edu (E.F. Walker), BCornblatt@northwell.edu (B.A. Cornblatt), Gary.Brucato@nyspi.columbia.edu (G. Brucato), spurcell@telesage.com (S.E. Purcell), brodey@email.unc.edu (I.S. Brodey), kcadenhead@ucsd.edu (K.S. Cadenhead).

https://doi.org/10.1016/j.schres.2019.01.015

The Psychosis Risk Calculator (Cannon et al., 2016) combines SIPS results with those of several additional assessments to improve the predictive model, but the combined assessments take several hours to administer. The Prodromal Questionnaire instruments including the Brief Version (PQ-B) were developed as screening instruments to identify CHR individuals (Loewy et al., 2005, 2011a). Although other instruments have been developed for screening purposes, the PQ-B is the most well established self-report screener for CHR status (Jarrett et al., 2012; Kline et al., 2012a, 2012b; Loewy et al., 2011b; Okewole et al., 2015). The measure yields a high false-positive rate for CHR status, however, which may render it unsuitable for widespread use as a screener in general populations (Kline et al., 2012b; Xu et al., 2016). Fusar-Poli et al. (2017) have also developed a model that uses demographic variables and current diagnosis to predict progression to psychosis, but the model is not specific to early psychosis or schizophrenia. Kobayashi et al. (2008) validated the PRIME Screen-Revised (PS-R) against 6-month conversion in a Japanese population, but only four people in their study actually converted. Given the low prevalence of early psychosis and the resources required for a clinician-administered assessment, it is desirable to have a highly specific self-report screener for early psychosis to promote appropriate early intervention (Cohen and Marino, 2013; Comparelli et al., 2014).

In an earlier project (Brodey et al., 2018a), we developed a self-report item bank to serve as the foundation for developing an early psychosis screener. 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 iterative rounds of cognitive interviewing per item (Willis, 2005). We identified a subset of 148 items that were well understood by CHR individuals and public sector mental health care recipients, and that our expert panel believed would cover the breadth of concepts associated with the prodromal period and early psychosis. In further analysis of this sample, we used Spectral Clustering and Minimum Redundancy Maximum Relevance procedures to select a subset of self-report items, the Early Psychosis Screener (EPS)-26, that could differentiate Clinically Low Risk (CLR) and CHR individuals with excellent accuracy (ROC AUC = 0.899), although accuracy for discriminating CHR from first episode psychosis (FEP) was lower (AUC = 0.614) (Brodey et al., 2018b). A printable copy of the complete EPS-26 assessment can be found at www.eps.telesage.org.

In initiating the present study, we wanted to validate a screening instrument against *true psychotic conversion* using a patient sample gathered from established CHR clinics affiliated with the North American Prodrome Longitudinal Study (NAPLS) and Center of Prevention and Evaluation (COPE, New York State Psychiatric Institute). We chose our analytic methods a priori, before any exploration of possible predictive models for actual psychotic conversion. In developing this plan, our initial goal was to develop an assessment that would identify individuals in need of in-person evaluation and possible treatment. Our plan included a Spectral Clustering Analysis followed by Support Vector Machine (SVM) development and testing.

Our hypothesis was that Spectral Clustering Analysis followed by a nonlinear SVM analysis could be used to put together a subset of the 148 self-report items that would accurately identify people who were at high risk of experiencing a psychotic disorder within 12 months and/or who had already experienced a first episode psychosis (FEP). Whereas the norm has been to validate against a proxy outcome such as CHR status, we hoped that validating against conversion or FEP status would result in a more useful assessment.

2. Methods

2.1. Participants

Recruitment was conducted at eight clinical sites: Calgary, New York State Psychiatric Institute, Emory, UCLA, UCSD, UNC-CH, Yale, and Zucker Hillside Hospital. TeleSage served as the only non-clinical recruitment site. All participants provided IRB-approved informed consent. Exclusionary criteria included: presence of a medical condition known to affect the central nervous system, estimated IQ <70, and age <14 or >35. The recruitment procedures for the NAPLS sites and COPE were based on the Criteria of Psychosis-risk Syndromes (COPS), which are delineated in the SIPS, and have been comprehensively described in the literature (Addington et al., 2012; Brucato et al., 2017).

Clinical participants in this study were recruited from a pool of patients referred to a NAPLS research site or the COPE research clinic for evaluation of psychosis risk. All participants were already receiving a SIPS as part of their evaluation for the primary study (see Miller et al., 2003 for a description of the SIPS assessment procedures) and were asked to participate in the Early Psychosis for Screener for Internet (EPSI) study. For this study, the SIPS scores were used as described by McGlashan et al. (2001) to divide help-seeking participants into three eligible groups: CLR, CHR, and FEP. (Please note that attenuated positive symptoms better accounted for by another psychiatric condition, as assessed using the SCID, represented an exclusionary criterion.) CLR participants were those who were referred to the specialty clinic, but who were found to have a SIPS P score of 1 or 2. All participants completed paper assessments including demographic items, our 148 test items, and the PQ-B.

We also recruited an independent sample of non-clinical control participants in Chapel Hill, North Carolina. These individuals were half-time or greater students, between the ages of 18 and 35, with no history of mental health treatment during the previous 2 years. (We did not conduct a structured interview to confirm an absence of CHR or FEP in the control population.) These individuals were not followed longitudinally.

2.2. Items

We began the study by administering the 148 test items. Afterwards, we removed items that might lead to spurious differentiations for reasons unrelated to the true likelihood of 12-month conversion. For this reason, we removed the demographics items. We also removed all data on participants' alcohol and drug use from the algorithm development process. Drug usage varied greatly and no particular drug other than marijuana was regularly endorsed. We also considered the potential inaccuracy of self-report drug use data. Finally, we removed items which were not applicable to all participants (i.e., specific work- and study-related items). In all, we removed 24 items from the analysis a priori, leaving 124 test items. Although it is certainly possible that some of these 24 items would have yielded useful information, given our limited sample size, we did not want to risk the need for post hoc analysis.

2.3. Spectral clustering analysis

A detailed description of the spectral clustering techniques employed in this study can be found in the online supplement (Shi and Malik, 2000; Ng et al., 2001; von Luxburg, 2007).

2.4. Support vector machine classifiers

SVM classifiers were trained to predict, given an individual's responses to the EPSI items, whether this individual would meet diagnostic criteria for a psychotic disorder within 12 months or was already experiencing an FEP. SVMs are a versatile class of supervised machine-learning methods that can be trained to learn linear or nonlinear input-output relations (Schölkopf and Smola, 2002; Vapnik, 1998). SVMs do not suffer much from the local minima problem of getting stuck in a suboptimal solution of the task. They have excellent generalization abilities and thus a reduced likelihood of overfitting. Here we

used the SVM^{light} software package (available for download at http://svmlight.joachims.org/) (Joachims, 1999).

Our a priori chosen approach was to use the bootstrap aggregating method (Breiman, 1996) to develop 50 radial basis function (RBF)-kernel SVM classifiers, each trained on 80% of all the subjects, selected randomly with replacement from the entire pool. Once developed, these 50 classifiers were used together as a battery to classify any given participant by averaging their scores, assigning him or her to either of the two discriminated groups based on whether the average battery score was above or below the classification threshold (default = 0). Only default SVM parameters were used, except for the *gamma* parameter, which specifies the RBF variance. Based on our earlier SVM studies of early psychosis item pool responses, we set *gamma* = 0.01. Cross-validation was done using the standard leave-one-out approach. A more detailed description of the SVM training and testing techniques employed in this study can be found in the online supplement.

The choices of (1) using 50 classifiers, (2) training each of them on 80% of subjects, and (3) using nonlinear SVMs with RBF kernel whose parameter *gamma* was set to 0.01, were made a priori, based on general considerations, in order to avoid inflating the classifier performance and to allow cross-validation using the leave-one-out approach. Several post-hoc models were also explored.

3. Results

3.1. Participants

We recruited a total of 353 participants from the eight early psychosis sites where SIPS screening is part of the standard protocol. Six of the clinical participants had missing intake data. (In most cases they had missed one or more pages of the test items.) The response sets from these participants were excluded from further analysis, leaving 347 clinical response sets. These can be further divided into: 71 CLR, 234 CHR, and 42 FEP. We also recruited 107 students from Chapel Hill, North Carolina to serve as non-clinical control (NCC) participants. Overall, participants were 20.5 ± 4.4 years old and 41% males.

We attempted to follow all 234 CHR participants longitudinally for a minimum of 12 months and for up to 24 months. Eighty-five of these participants were lost to follow-up at the various sites either prior to converting or prior to the end of the initial 12-month period. The remaining 149 participants were followed either to the time of conversion or for a minimum of 12 months. Thirty-four of the 149 CHR participants converted during the first 12 months, and 5 participants converted between 12 and 24 months. Seventy-five did not convert during the first 12 months, but were lost to follow-up prior to 24 months. Thirty-five participants had not converted after 24 months. In summary, 34 participants converted to psychosis within one year and 115 participants are known not to have converted during the first 12 months.

For the purpose of training SVMs, participants were combined into two groups. The non-converter group, Group *NC*, included CLR participants (n = 71), as well as CHR participants who did not convert even after 2 years (n = 35), for a total of 106 participants. We selected this population for SVM training to ensure that we would train SVMs with a broad group of participants who were referred for evaluation at a specialty clinic, but who were found to be in a low risk group or not to have converted during the longest possible follow-up period. The converter group, Group *C*, included FEP participants (n = 42) as well as those CHR participants who converted within the first 12 months after their initial SIPS screening (n = 34), for a total of 76 participants. Training SVMs with these groups would minimize the likelihood that we would misclassify participants who might benefit from an in-person evaluation and served to maximize the public health relevance of the ultimate EPSI.

The CHR participants who converted between 12 and 24 months after initial screening (n = 5) were not used in training SVMs as there were too few for us to be able to use them to develop algorithms to predict 24-month conversion. CHR participants who did not convert during the first 12 months and who dropped out of the study between 12 and 24 months after intake (n = 75) were also excluded from the SVM classifier training as a small number of these participants might have converted between 12 and 24 months after intake. Data from these individuals were subsequently used as part of an independent validation of the algorithms.

3.2. Item reduction using spectral clustering analysis (SCA)

When SCA was performed, the 124 items formed two distinct clusters. An overwhelming majority of the 64 positive items target either psychosis or mania. In contrast, the 60 negative items predominantly target depression, anxiety, and social and general work/school functioning. We therefore confined our SVM study exclusively to the 64 positive items. It should be pointed out that our spectral clustering procedure did not rely on the membership of the individuals in the *C* and *NC* groups to partition the 124 items. In the current study, we did not have sufficient power to reduce the 64 items relating to positive symptoms or to integrate the 60 negative symptom items into our algorithm development process. Please see the online supplement for a more detailed description of the SCA results.

3.3. EPSI discriminative performance

The 64 items of the positive spectral cluster were used as inputs to a bootstrap aggregating battery of 50 SVMs, trained to classify any given individual as belonging to either Group *C* or Group *NC*. Each SVM was trained on 80% of all the subjects in Groups *C* and *NC*, selected randomly with replacement from the entire pool after one subject was removed and saved for leave-one-out cross-validation testing.

The cross-validated Group *C* vs. Group *NC* performance of the 50-SVM battery is listed in Table 1. For a comparison, Table 1 also lists the same classification performance of other screening instruments; i.e., SIPS, PQ-B at the published sum score cut-offs of 3 and 6, and EPS-26 at the published sum score cut-off of 33.5. Although CLR subjects in this project were not followed to determine whether they converted after the initial screening, in Table 1 we used an existing estimate that 1.8% of those individuals would have converted (Webb et al., 2015). Table 1 shows that the EPSI is superior to the SIPS, PQ-B, and EPS-26 in its avoidance of false positives. Access to a full list of the EPSI items and to the self-scoring online EPSI can be found at www.eps.telesage.org.

In addition to its cross-validation testing on Group *C* and Group *NC* individuals, EPSI was also applied to the independent sample of 107 NCC individuals, as well as 5 CHR individuals who did not convert during the first 12 months but did convert during the subsequent 12 months, and 75 CHR individuals who did not convert during the first 12 months, but were lost to follow up prior to 24 months (the last 2 categories are CHR definite 1-year non-converters). Table 2 lists the numbers of individuals in each category (non-clinical control, CLR,

Table 1
Classification performance of SIPS, PQ-B, EPS-26 and EPSI screeners in correctly distinguishing between FEP/converter subjects (Group *C*) and non-converter subjects (Group *NC*). The PQ-B was scored with published sum score cutoffs of 3 and 6 (Loewy et al., 2011a). The EPS-26 was also scored at its published sum score cutoff of 33.5. *True Positive Fraction* – probability that a Group *C* subject will be correctly classified as such. *False Positive Fraction* – probability that a Group *NC* subject will be misclassified as Group *C*. *Positive Predictive Value* – probability that a subject classified as Group *C* does indeed belong to Group *C*.

Method	SIPS	PQ-B ≥3	PQ-B ≥6	EPS-26 ≥33.5	EPSI	EPSI+ SIPS
True Positive Fraction (%)	98.2	84.2	71.1	81.6	51.3	76.3
False Positive Fraction (%)	33.0	51.9	37.7	45.3	11.3	8.5
Positive Predictive Value (%)	68.5	53.8	57.5	56.4	76.5	86.6

Table 2
Percent (number) of participants in each category who were classified as belonging to Group *C* (i.e., FEP/converter subjects). * marks categories that were not used in SVM training.

Method	PQ-B ≥3	PQ-B ≥6	EPS-26 ≥33.5	EPSI
Non-Clinical* (n = 107)	27.1 (29)	12.2 (13)	29.0 (19)	0.9 (1)
CLR (n = 71)	29.6 (21)	16.9 (12)	19.7 (14)	4.2 (3)
CHR-NC (n = 35) 2-year non-converters	97.1 (34)	80.0 (28)	97.1 (34)	25.7 (9)
CHR-NC* (n = 80) 1-year non-converters	92.5 (74)	72.5 (58)	83.8 (67)	21.3 (17)
CHR-C (n = 34) 1-year converters	85.3 (29)	70.6 (24)	79.4 (27)	47.1 (16)
FEP (n = 42)	83.3 (35)	71.4 (30)	83.3 (35)	54.8 (23)

CHR definite 2-year non-converters, CHR definite 1-year non-converters, CHR 1-year converters, and FEP) who were classified as belonging to Group *C* by each of the studied screening instruments. It shows that EPSI had a lower probability of mistakenly classifying non-clinical, CLR, or CHR non-converter individuals as converters than other methods. Also note that EPSI classification performance on the individuals in the non-clinical and CHR 1-year non-converter categories, which were not used in SVM training, agrees closely with its cross-validation performance on the individuals in the CLR and CHR 2-year non-converter categories, respectively. However, the false negative rate of classification was also lower than with the other screeners in the FEP subjects and in the SIPS CHR converters.

Focusing specifically on the ability of the studied screeners to predict who, among CHR individuals in the present dataset, will convert within one year, we list their positive predictive values (PPV), sensitivity, and specificity in Table 3. (For Table 3 we did not include any CLR or FEP participants.) Since Table 3 relies exclusively on participants who were found to be CHR on the SIPS, *the results reflect the performance statistics of the SIPS alone or in combination with each of the instruments noted. Again, we find that the PPV and specificity of the EPSI + SIPS was superior to the SIPS alone, or PQ-B + SIPS, or EPS-26 + SIPS. The PQ-B did not appear to add information when administered in conjunction with the* SIPS.

Our battery of 50 SVMs classifies any given individual based on his or her average battery score. This battery score can be viewed as an SVM-learned estimation of a person's position on a "psychotic thinking" spectrum. This spectrum is shown in Fig. 1, which plots the battery-averaged EPSI scores of all studied groups. In the figure, information on 30-day drug and alcohol use was re-introduced, and the 107 non-clinical controls were divided into two groups. The first group comprised the 60 participants who denied using any drugs other than alcohol to get high during the previous 30 days. The second group of 47 participants reported that they had used drugs at least once in the previous 30 days to get high.

After performing the analyses selected a priori, we explored several post-hoc analyses. These included, for example, using all 124 items, but condensing the items into 30 variables using PCA. We also explored varying model parameters and using a linear SVM model. We explored training the algorithm with different sub-populations, e.g. only CHR individuals using a simple outcome variable of conversion vs. non-conversion. These models tended to converge on similar results. Individual models were somewhat superior with regard to specific purposes, but none of the additional models tested were clearly superior overall.

Table 3
Statistical performance of the SIPS alone or a combination of the SIPS with (a) the PQ-B with the published cutoffs of 3 and 6, or (b) the ESP-26 with a published cutoff of 33,5, or (c) the EPSI in predicting actual conversion of CHR participants within 12 months.

Method	SIPS	PQ-B ≥3	PQ-B ≥6	EPS-26 ≥33.5	EPSI
Positive Predictive Value (%)	22.8	21.2	21.8	21.1	38.1
Sensitivity (%)	100	85.3	70.6	79.4	47.1
Specificity (%)	0	6.1	25.2	12.2	77.4

4. Discussion

We wish to emphasize that the analytic techniques employed in this study were selected a priori. SCA enabled us to identify a useful and cohesive subset of items in order to improve the item-to-participant ratio. Since the SCA technique employed did not take group assignment into consideration, the item reduction process did enhance the predictive capacity of the items relative to the item group as a whole. Although it is possible that other machine learning strategies might have yielded superior prediction, we considered nonlinear SVM to be among the best a priori strategies for developing a predictive algorithm. We did not test any other algorithms.

Overall, the PPV for the EPSI at the default threshold was superior to that of the SIPS, the EPS-26, and the PQ-B. Furthermore, while the SIPS is a loosely structured interview that requires substantial training and about 90 min to administer, *the EPSI is a 64-item self-report assessment that is intended to have a 5th-grade reading level and take <12 min to administer*. The PQ-B is the primary alternative self-report early psychosis screener; it was well ahead of its time when it was first released with 96 items (Loewy et al., 2005). It is now comparatively short at 21 items, yet even at the more specific upper published cut-off of 6 it is probably too sensitive and not specific enough to be used for screening either a general population or individuals seeking an evaluation for early psychosis. The EPS-26, developed as an earlier part of this project, was validated against SIPS status and may provide a simple strategy for identifying individuals who would score in the CHR range on the SIPS, but it was not validated against true conversion and is considerably less accurate than the EPSI at predicting true conversion. Just as the length of the PQ-96 was reduced when additional data became available, we expect that

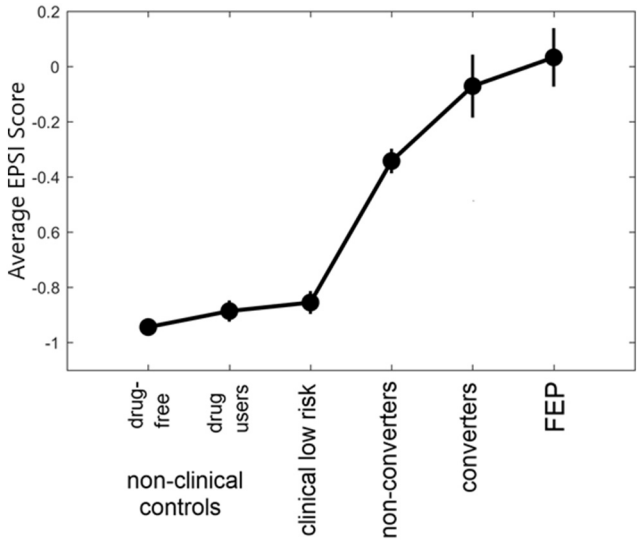

Fig. 1. Average EPSI scores for six studied groups: (1) NCCs who denied any 30-day drug use, (2) NCCs reporting some 30-day drug use, (3) CLR, (4) 12-month non-converters, (5) psychotic converters, and (6) FEP participants. The error bars are standard error of the mean. (SVM scores express relative proximity of tested data points to the classification boundary.)

the length of the 64-item EPSI will also decrease over time. Since items with low predictive value generate noise in SVM algorithms, a shorter version of the EPSI is likely to be more accurate than the longer version. *The EPSI is the first self-report assessment that can be used effectively to screen and refer large numbers of people, who are thought by friends, relatives, or primary care clinicians to be at increased risk of psychosis, for in-person evaluation by a licensed professional. Due to its high positive predictive value, the EPSI can serve as the wide end of a funnel that brings high-risk people in for evaluation as part of routine clinical care or a specific research protocol.* Because of it uses machine leaning scoring algorithms, administration over the internet via a smartphone or laptop is ideal.

The EPSI may also have an important role at the narrower end of the funnel. In our study, only 23% of all CHR individuals actually converted. In a clinical trial this means that if all CHR individuals are randomized, 77% of the participants in both the active and control arms have only a minimal potential to improve, even with a highly efficacious intervention, since they do not have the condition. In this situation, the beneficial effects of the intervention are unlikely to be recognized; however, if the SIPS and EPSI have a combined ability to identify a population of CHR individuals, 38% of whom will convert, as they did in our study, then it will be easier to identify an effective intervention.

There is an added benefit to using the EPSI. Because it characterizes the severity of granular symptoms on a Likert scale, it is possible to identify the specific symptoms that improve as a result of an intervention. This is especially important because we recognize that schizophrenia is a heterogeneous set of disorders. By identifying specific symptom clusters that are responsive to an intervention, the EPSI should in the future further improve the signal-to-noise ratio in clinical trials. This will further increase the likelihood that an efficacious intervention will be identified. Just as SVM classifiers can be used to assess the combined risk of specific patterns of alleles in an individual, it should also be possible to use the SVM classifiers to dissect associated patterns of symptoms.

This study used the leave-one-out cross-validation model, but it is noteworthy that the non-clinical sample was entirely independent. Of these 107 participants, who were recruited at a separate site and whose data were NOT used in algorithm development, <1% of participants were identified as possible converters. Similarly, data from the 80 CHR individuals in the 12-month non-converter sample were NOT used in algorithm development. But, among these 80 individuals, only 21% screened positive on the EPSI. This finding is fully consistent with and tends to confirm our results.

5. Limitations

A marked limitation of the EPSI is its low sensitivity at recognizing FEP or future converters in the CLR + CHR population: at its default SVM classification threshold of '0', it has a low sensitivity of 0.5 although a high specificity of 0.89. At this threshold, for every person identified with an FEP or as being a future CHR converter, one person will be missed. Sensitivity can be improved by reducing the SVM classification threshold to, for example, −0.7 (Fig. 2). This yields a sensitivity of 0.78 and a specificity of 0.58. Still, these sensitivities and specificities, while within the range of many established medical screeners, e.g. the Fasting Plasma Glucose Test for gestational diabetes (Maxim et al., 2014), are not high enough to consider the EPSI a diagnostic tool. Any use of the EPSI must take into consideration this high false negative rate. Our hope is that the EPSI will be used as a screener to identify individuals who should be referred to specialty providers for further in-person evaluation for prodromal status. Using the EPSI sub-population curves presented in Fig. 2 and by estimating the true proportion of 'non-clinical control', 'CLR', 'CHR non-converter', 'CHR converter', and 'FEP' individuals in any population to be evaluated, the reader can generate the anticipated EPSI screening results, PPV and NPV. An ROC curve for the EPSI is provided in the online supplement, Fig. S2. In determining the EPSI's clinical utility for any given population, one must consider

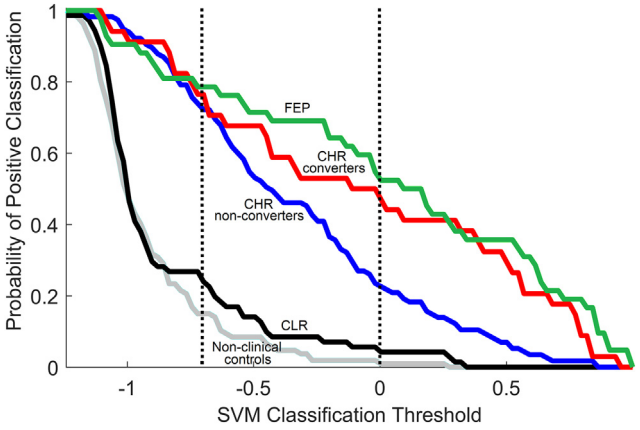

Fig. 2. Probability that an individual taken from either of the studied groups will be classified by the EPSI as belonging to Group *C* (FEP/converter) plotted as a function of the SVM classification threshold. By plotting a vertical line anywhere along the x axis, it is possible to assess the impact of changing the threshold on classifications within each sub-group and to model how the EPSI will do in diverse screening situations.

these metrics as well as the purpose of the investigation. In the hypothetical scenarios described below, we have set the EPSI cut-off either at its default SVM classification threshold of 0, at a somewhat more sensitive threshold of −0.7, or at a fairly specific threshold of +0.4. One solution to the low sensitivity of the EPSI would be to have all participants who screen between '-0.7' and '0' on the EPSI rescreened after 8 weeks. While this option is still under study, rescreening or measuring progression may ultimately improve the sensitivity of the screening system.

Scenario 1: A clinic selectively screens new clients between the ages of 14 and 35 who seem to have unusual thoughts or perceptions, or who exhibit social withdrawal. We assume that 5% are experiencing an FEP, 5% will be 'CHR converters', and 15% will be 'CHR non-converters'. The remaining 75% of the sample are CLR. For this enriched sample, we want high sensitivity. At a cut off of "-0.7", for every 100 people screened we will identify 8 true positives, 29 false positives, 2 false negatives, and 61 true negatives. For every CHR Converter that we identify, we will identify 2.75 false positives from among the CHR non-converters, a ratio that is close to that of the SIPS. The screening ratio for true positives and false positives is 28%.

Scenario 2: The same clinic decides to screen all new clients using the EPSI. About 1% of the total sample has FEP, 1.25% of the individuals screened will be CHR-converters, and 3.75% will be CHR non-converters. The remaining 94% are somewhere between CLR and NCCs. For this scenario we will assume that they are CLRs, as this is the more challenging screening scenario. Based on this scenario and the curves shown in Fig. 2 for the EPSI, and using a cut off of "0" to promote specificity, if we screen 100 individuals we will identify 1 true positive, 1 false negative, 5 false positives, and 93 true negatives. Thus, for every 6 people sent for in-person screening with the SIPS, we will identify 1 true positive and 5 true negatives. For every CHR converter we identify we will identify 1 false positive from among the CHR non-converters, a ratio that is much better than that of the SIPS.

Scenario 3: An investigator who wishes to screen a very large number of potential participants quickly for an early psychosis or FEP study might want to use the online assessment to screen help-seeking individuals between the ages of 18 to 25. Let's assume that among each 1000 people assessed, 5 have FEP, 5 are CHR-converters, 15 are CHR non-converters, 200 are CLRs, and 775 are non-clinical controls. In order to avoid a high false positive rate, he or she might want to increase the threshold to +0.4. This would still enable the EPSI to identify about 36% of FEP and CHR converters. It would also identify about 10% of CHR non-converters, a majority of whom would go on to develop other diagnosable disorders. The screening would miss approximately 64% of the FEP and CHR converter populations. However, at this high stringency,

none of the non-CHR individuals would be identified as false positives. In certain situations, high-volume, low-cost screening in somewhat enriched populations could prove to be a cost-effective screening strategy to promote early identification or recruitment.

The second potential limitation of this study would be overestimation of the accuracy of the assessment. As described above, we took many precautions to prevent inflation of the accuracy of the EPSI; however, repeating the evaluation with a complete fully independent sample is the only sure way to demonstrate the accuracy of the assessment.

While the high specificity of the EPSI may make it an excellent screener for use in making specialty referrals, it is not a diagnostic tool. People can respond to self-report assessments in a variety of ways for many complex reasons. For example, among the FEPs there were several participants who endorsed very few symptoms, despite being rated SIPS six by the interviewer. Our hypothesis is that these participants minimized their symptoms on the self-report assessment, but that the interviewer was able to elicit the underlying symptoms. To improve the accuracy of the EPSI, and with the help of NAPLS and COPE investigators, we developed a few items to identify people who tend to 'minimize' or 'emphasize' the severity of their symptoms. In the future, this short scale should provide independent information that may increase the accuracy of the EPSI. At present we believe that only a specialist should make a diagnosis of APS or FEP.

An additional limitation of the EPSI is that it was not able to distinguish effectively between prodromal and FEP states. This suggests that it is difficult to demarcate an exact line separating the later prodromal and early FEP periods, most likely because the distinction rests on the client's 'conviction' regarding the reality of his or her delusions or hallucinations. "Conviction" itself is probably best viewed as an episodic trait rather than a single binary variable.

In the future, since attenuated psychotic syndrome (APS) and perhaps even FEP represent a progressive continuum of symptoms, we expect that a longitudinal use of the EPSI may be most clinically useful: it is likely that the progression of EPSI scores over several administrations might prove more predictive of conversion than individual EPSI scores at any single point in time. Multi-step screening, using items relating to negative symptoms, might also increase overall prediction accuracy. The EPSI may also prove useful for tracking the progression of APS in clinical trials.

6. Conclusions

The EPSI is the first self-report assessment validated against 12-month conversion that can be used effectively to screen and refer large numbers of people who are thought by friends, relatives, or primary care clinicians to be at increased risk of psychosis. The EPSI is not a diagnostic tool, but it currently has the potential to be used for public health screening, both to decrease the duration of untreated psychosis and to bring eligible individuals into specialty clinical care settings and clinical trials aimed at identifying more effective treatments for early psychosis. By combining the EPSI with the SIPS, we can further increase the PPV of each individual assessment. Despite its strengths, the implications of the false negative rate must be taken into consideration in any specific screening scenario. The cut-off for the EPSI can easily be modified to increase or decrease sensitivity of the assessment based on purpose.

Role of the funding source

This study was supported by a National Institute of Mental Health (NIMH) grant #MH094023. The funding source did not have a role in the study design.

Contributors

Dr. Benjamin Brodey designed and coordinated the study. Dr. Ragy Girgis was the site PI at the New York State Psychiatric Institute at Columbia University Medical Center. Dr. Oleg Favorov analyzed the data and wrote the Methods and Results sections of the manuscript. Dr. Jean Addington was the site PI at the University of Calgary. Dr. Diana Perkins was the site PI at the University of North Carolina at Chapel Hill. Dr. Carrie Bearden was the site PI at the University of California Los Angeles. Dr. Scott Woods was the site PI at the PRIME Psychosis Prodrome Research Clinic and assisted with the design of the study. Dr. Elaine Walker was the site PI at Emory University. Dr. Barbara Cornblatt was the site PI at The Zucker Hillside Hospital. Dr. Brucato assisted with participant recruitment and assessment at Columbia University. Dr. Susan Purcell assisted with preparation of the data set and the manuscript. Dr. Inger Brodey assisted with the editing of the manuscript and consulted on the development of the methodology and documentation. Dr. Cadenhead was the site PI for the University of California San Diego. All authors contributed to and have approved the final manuscript.

Conflict of interest

Drs. Inger and Benjamin Brodey are co-owners of TeleSage, Inc. Dr. Ragy Girgis received research support from Genetech, Otsuka Pharmaceutical, Allergan, and BioAdvantex Pharma. Dr. Susan Purcell is the Senior Research Associate at TeleSage, Inc. All other authors declare that they have no conflicts of interest.

Acknowledgement

Warren Brodey MD, psychiatrist and cyberneticist, has been teaching the corresponding author about densely interconnected non-linear systems since 1959.

Funding for this project was made possible by the National Institute for Mental Health (NIMH), Grant #MH094023.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.schres.2019.01.015.

References

Addington, J., Cadenhead, K.S., Cornblatt, B.A., et al., 2012. North American Prodrome Longitudinal Study (NAPLS 2): overview and recruitment. Schizophr. Res. 142, 77–82. https://doi.org/10.1016/j.schres.2012.09.012.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140.

Brodey, B.B., Addington, J., First, M.B., et al., 2018a. The Early Psychosis Screener (EPS): item development and qualitative validation. Schizophr. Res. 197, 504–508. https://doi.org/10.1016/j.schres.2017.11.027.

Brodey, B.B., Girgis, R.R., Favorov, O.V., et al., 2018b. The Early Psychosis Screener (EPS): Quantitative validation against the SIPS using machine learning. Schizophr. Res. 197, 516–521. https://doi.org/10.1016/j.schres.2017.11.030.

Brucato, G., Masucci, L., Arndt, Y., et al., 2017. Baseline demographics, clinical features and predictors of conversion among 200 individuals in a longitudinal prospective psychosis-risk cohort. Psychol. Med. 47 (11), 1923–1935. https://doi.org/10.1017/S0033291717000319.

Cannon, T.D., Yu, C., Addington, J., et al., 2016 Oct 1. An individualized risk calculator for research in prodromal psychosis. Am. J. Psychiatry 173 (10), 980–988 (Epub 2016 Jul 1). https://doi.org/10.1176/appi.ajp.2016.15070890.

Cohen, C.I., Marino, L., 2013. Racial and ethnic differences in the prevalence of psychotic symptoms in the general population. Psychiatr. Serv. 64 (11), 1103–1109. https://doi.org/10.1176/appi.ps.201200348.

Comparelli, A., De Carolis, A., Emili, E., et al., 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. https://doi.org/10.1016/j.comppsych.2014.01.006.

Fusar-Poli, P., Rutigliano, G., Stahl, D., Davies, C., Bonoldi, I., Reily, T., McGuire, P., 2017. Development and validation of a clinically based risk calculator to the transdiagnostic prediction of psychosis. JAMA Psychiat. 74 (5), 493–500. https://doi.org/10.1001/jamapsychiatry.2017.0284.

Jarrett, M., Craig, T., Parrott, J., et al., 2012. Identifying men at ultra high risk of psychosis in a prison population. Schizophr. Res. 136 (1–3), 1–6. https://doi.org/10.1016/j.schres.2012.01.025.

Joachims, T., 1999. Making large-scale SVM learning practical. In: Schölkopf, B., Burges, C., Smola, A. (Eds.), Advances in Kernel Methods - Support Vector Learning. MIT-Press, Cambridge, MA.

Kline, E., Schiffman, J., 2014. Psychosis risk screening: a systematic review. Schizophr. Res. 158, 11–18. https://doi.org/10.1016/j.schres.2014.06.036.

Kline, E., Wilson, C., Ereshefsky, S., et al., 2012a. Convergent and discriminant validity of attenuated psychosis screening tools. Schizophr. Res. 134 (1), 49–53. https://doi.org/10.1016/j.schres.2011.10.001.

Kline, E., Wilson, C., Ereshefsky, S., et al., 2012b. Psychosis risk screening in youth: a validation study of three self-report measures of attenuated psychosis symptoms. Schizophr. Res. 141 (1), 72–77. https://doi.org/10.1016/j.schres.2012.07.022.

Kobayashi, H., Nemoto, T., Koshikawa, H., Osono, Y., Yamazawa, R., Murakami, M., Kashima, H., Mizuno, M., 2008. A self-reported instrument for prodromal symptoms of psychosis: testing the clinical validity of the PRIME Screen—Revised (PS-R) in a Japanese population. Schizophr. Res. 106 (2–3), 356–362. https://doi.org/10.1016/j.schres.2008.08.018.

Loewy, R.L., Bearden, C.E., Johnson, J.K., Raine, A., Cannon, T.D., 2005. The prodromal questionnaire (PQ): preliminary validation of a self-report screening measure for prodromal and psychotic syndromes. Schizophr. Res. 79, 117–125. https://doi.org/10.1016/j.schres.2005.03.007.

Loewy, R.L., Pearson, R., Vinogradov, S., Bearden, C.E., Cannon, T.D., 2011a. Psychosis risk screening with the prodromal questionnaire – brief version (PQ-B). Schizophr. Res. 129, 42–49. https://doi.org/10.1016/j.schres.2011.03.029.

Loewy, R.L., Therman, S., Manninen, M., Huttunen, M.O., Cannon, T.D., 2011b. Prodromal psychosis screening in adolescent psychiatry clinics. Early Interv. Psychiatry 6, 69–75. https://doi.org/10.1111/j.1751-7893.2011.00286.x.

Maxim, L.D., Niebo, R., Utell, M.J., 2014. Screening tests: a review with examples. Inhal. Toxicol. 26 (13), 811–828.

McGlashan, T.H., Miller, T.J., Woods, S.W., Hoffman, R.E., Davidson, L., 2001. A scale for the assessment of prodromal symptoms and states. In: Miller, T.J., Mednick, S.A., McGlashan, T.T., Liberger, J., Johannessen, J.O. (Eds.), Early Intervention in Psychotic Disorders. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 135–149.

Miller, T.J., McGlashan, T.H., Woods, S.W., et al., 1999. Symptom assessment in schizophrenic prodromal states. Psychiatry Q. 70 (4), 273–287. https://doi.org/10.1023/A:1022034115078.

Miller, T.J., McGlashan, T.H., Rosen, J.L., et al., 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. https://doi.org/10.1176/appi.ajp.159.5.863.

Miller, T.J., McGlashan, T.H., Rosen, J.L., et al., 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. https://doi.org/10.1093/oxfordjournals.schbul.a007040.

Ng, A.Y., Jordan, M.I., Weiss, Y., 2001. On spectral clustering: analysis and an algorithm. In: Dietterich, T.G., Becker, S., Ghahramani, Z. (Eds.), Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic. MIT Press, Cambridge, MA, pp. 849–856.

Okewole, A.O., Ajogbon, D., Adenjii, A.A., et al., 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. https://doi.org/10.1016/j.schres.2015.01.006.

Schölkopf, B., Smola, A.J., 2002. Learning with Kernels. MIT Press, Cambridge, MA.

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

Vapnik, V., 1998. Statistical Learning Theory. Wiley, New York.

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

Webb, J.R., Addington, J., Perkins, D.O., et al., 2015. Specificity of incident diagnostic outcomes in patients at clinical high risk for psychosis. Schizophr. Bull. 41 (5), 1066–1075. https://doi.org/10.1093/schbul/sbv091.

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

Xu, L.H., Zhang, T.H., Zheng, L.N., et al., 2016. Psychometric properties of prodromal questionnaire – brief version among Chinese help-seeking individuals. PLoS One https://doi.org/10.1371/journal.pone.0148935.