Schizophrenia Research xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Schizophrenia Research

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

The Early Psychosis Screener (EPS): Item development and qualitative validation

Brodey B.B. [a,*], Addington J. [b], First M.B. [c], Perkins D.O. [d], Woods S.W. [e], Walker E.F. [f], Walsh B. [e], Nieri J.M. [d], Nunn M.B. [g], Putz J. [h], Brodey I.S. [a,i]

[a] TeleSage, Inc., 201 East Rosemary St., Chapel Hill, NC 27514, USA
[b] Hotchkiss Brain Institute, Department of Psychiatry, University of Calgary, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6, Canada
[c] Department of Psychiatry, Columbia University Medical Center, 1051 Riverside Drive, New York, NY 10032, USA
[d] Department of Psychiatry, School of Medicine, The University of North Carolina at Chapel Hill, 101 Manning Dr, Chapel Hill, NC 27514, USA
[e] PRIME Psychosis Prodrome Research Clinic, Connecticut Mental Health Center, B-38, 34 Park Street, New Haven, CT 06519, USA
[f] Departments of Psychology and Psychiatry, Emory University, 36 Eagle Row, Atlanta, GA 30322, USA
[g] Centerstone Tennessee, 1921 Ransom Place, Nashville, TN 37217, USA
[h] Centerstone Research Institute, 645 South Rogers Street, Bloomington, IN 47403, USA
[i] The University of North Carolina at Chapel Hill, 434 Greenlaw, Campus Box 3520, Chapel Hill, NC 27599, USA

ARTICLE INFO

Article history:
Received 27 July 2017
Received in revised form 6 November 2017
Accepted 24 November 2017
Available online xxxx

Keywords:
SIPS
PQ-B
CAARMS
NAPLS
Psychosis;
Prodromal
Schizophrenia
Screener

ABSTRACT

A panel of experts assembled and analyzed a comprehensive item bank from which a highly sensitive and specific early psychosis screener could be developed. Twenty well-established assessments relating to the prodromal stage, early psychosis, and psychosis were identified. Using DSM-5 criteria, we identified the core concepts represented by each of the items in each of the assessments. These granular core concepts were converted into a uniform set of 490 self-report items using a Likert scale and a 'past 30 days' time frame. Partial redundancy was allowed to assure adequate concept coverage. A panel of experts and TeleSage staff rated these items and eliminated 189 items, resulting in 301 items. The items were subjected to five rounds of cognitive interviewing with 16 individuals at clinically high risk for psychosis and 26 community mental health center patients. After each round, the expert panel iteratively reviewed, rated, revised, added, or deleted items to maximize clarity and centrality to the concept. As a result of the interviews, 36 items were revised, 52 items were added, and 205 items were deleted. By the last round of cognitive interviewing, all of the items were clearly understood by all participants. In future work, responses to the final set of 148 items and machine learning techniques will be used to quantitatively identify the subset of items that will best predict clinical high-risk status and conversion.

1. Introduction

Interest in identifying individuals at clinically high risk (CHR) of developing a psychotic spectrum disorder has grown over the past decade. The most widely used assessments are the Structured Interview for Psychosis-risk Syndromes (SIPS) and the Comprehensive Assessment of At-Risk Mental States (CAARMS) (Fusar-Poli et al., 2016). Both the SIPS and the CAARMS have very high sensitivity 91.6% (Webb et al., 2015; Fusar-Poli et al., 2016). Unfortunately, proper administration of these semi-structured interviews requires extensive training in order to assure high inter-rater reliability (Addington et al., 2012). Even with extensive SIPS or CAARMS training, only about 19.6% of individuals who are identified as CHR based on their SIPS score will actually go on to develop a psychotic disorder vs. 1.8% for help-seeking clinical controls. An additional 10.7% of CHR patients will develop bipolar disorder, unipolar depression, or an anxiety disorder vs. 11.8% of controls (Webb et al., 2015).

At present, the most widely used self-report screener for early psychosis is the Prodromal Questionnaire – Brief Version (PQ-B) (Loewy et al., 2005, 2011a). In general, the PQ instruments were extensively validated against the SIPS and the CAARMS (Loewy et al., 2011b; Ising et al.,

Abbreviations: CAARMS, Comprehensive Assessment of At-Risk Mental States; CHR, clinically high risk; CI, cognitive interviewing; IP, interviewer probe; NAPLS, North American Prodrome Longitudinal Study; PQ-B, Prodromal Questionnaire – Brief Version; PROMIS, Patient-Reported Outcomes Measurement Information System; SIPS, Structured Interview for Psychosis-risk Syndromes; TA, think aloud.

* Corresponding author.
E-mail addresses: bb@telesage.com (B.B. Brodey), jmadding@ucalgary.ca (J. Addington), mbf2@cumc.columbia.edu (M.B. First), diana_perkins@med.unc.edu (D.O. Perkins), scott.woods@yale.edu (S.W. Woods), psyefw@emory.edu (E.F. Walker), barbara.walsh@yale.edu (B. Walsh), jennifer_nieri@med.unc.edu (J.M. Nieri), brad.nunn@centerstone.org (M.B. Nunn), john.putz@centerstone.org (J. Putz), brodey@email.unc.edu (I.S. Brodey).

https://doi.org/10.1016/j.schres.2017.11.027
0920-9964/

2012). The PQ-B has high sensitivity, but as an outpatient screener it may lack sufficient specificity to make more widespread screening practical. There are a few additional difficulties with the PQ-B that this study aims to improve upon. First, although individual items are clearly written, they tend to focus on fairly mild symptoms that are common in the general population (e.g., seeing a fortune teller). Individual items also tend to combine several related but distinct experiences (e.g., "experiences with telepathy, psychic forces, or fortune telling") without any ability to distinguish between them. Finally, the primary response set for the PQ-B is a limited, binary 'yes/no'. A Likert scale was subsequently added, but its use is only indicated for items that are already endorsed with a 'yes'. The selective addition of a Likert scale also complicates statistical analyses.

Since the initial publication of the PQ-B in 2005, there have been several developments that can improve the creation of self-report screeners. These include cognitive interviewing (CI), which is useful in qualitative validation (DeWalt et al., 2007); Item Response Theory (IRT) and related aspects of Modern Measurement Theory (Reeve, 2002); machine learning strategies (Peng et al., 2005 & von Luxburg, 2007); and lessons learned from the National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS) initiative (Cella et al., 2007). While it is unusual for an item development manuscript not to conclude with a quantitative analysis defining the utility of the items, our premise has been that psychosis is one of the most difficult of human experiences to assess and that the quality of the items in an assessment naturally place an upper bound on predictive power, regardless of analytic strategy. (It is not possible to accurately and precisely interpret an item that is confusing, has multiple interpretations, or which includes several concepts e.g. in the case of depression: sad, depressed, or hopeless.) For these reasons, we have chosen to dedicate this manuscript to a detailed description of the application of these techniques to the development of a comprehensive self-report item bank that can be used to predict CHR status. The approach is primarily synthetic in nature, encompassing the theoretical frameworks for each of the assessments that form the basis of our item bank.

Our hypothesis is that we can develop a comprehensive set of simple Likert scale items that each represent a single, granular, core symptom associated with the prodromal period, including psychotic-like and psychotic experiences. Our belief is that this item bank will serve as the foundation for creating a self-report screener for early psychosis that could be used to predict SIPS CHR status and ultimately predict conversion with high specificity.

2. Methods

2.1. Stage I: item pool development

The first step was to gather widely used prodromal, early psychosis, and psychosis measures that have been described in the peer-reviewed literature. These measures are presented in a recent review of self-report and clinician-administered early psychosis screeners (Kline and Schiffman, 2014). Using these screeners and DSM-5 criteria, we identified the core concepts represented by each of the items in each of the assessments. These core concepts covered all of the criteria for schizophrenia spectrum and other psychotic disorders described in the DSM-5. Under the supervision of Dr. Brodey, who used similar techniques to develop the Perinatal Depression Inventory (Brodey et al., 2016), TeleSage staff rewrote items in a simplified self-report format. They based the items on a fifth-grade reading level, with one concept per item so that minimal interpretation of each item was required. Each item was intended to elicit a simple direct report of the individual's experiences and feelings. Wherever possible, items were written in a non-judgmental, non-pathologizing format. We avoided words and phrases with pejorative, multiple, or abstract connotations. (For an example, see the revised item 'I felt anxious.' in Results section, Table 1). Foreign words and words known to translate poorly into other languages were avoided.

Items were written to match a 5-point Likert scale (Never, Rarely, Sometimes, Often, Always) response set. This is the same response set that was used in the PROMIS initiative (DeWalt et al., 2007), except that we included a 'does not apply' response option as a second alternative to 'Never', for some items related to work and school experiences. This was for participants who were unemployed or not in school, therefore 'Never' could be ambiguous. We created items intended to represent different extremes of a symptom so as not to rely exclusively on the Likert scale for differentiation (Comparelli et al., 2014). Furthermore, we attempted to avoid items that might have a ceiling or floor effect. Items were written with a standard 'past thirty days' time frame. The panel of experts discussed using a scale of severity or distress instead of frequency; however, no single scale appeared to work perfectly to assess the prodromal period. Frequency appeared to act as an adequate proxy to capture intermittent prodromal episodes as well as attenuated symptoms. Although the PROMIS initiative used a 'past seven days' time frame, we reasoned that we needed a longer time frame in order to pick up the episodic symptoms that are associated with the prodromal period. Patients tend to answer consistently when asked about frequency or intensity, so the panel concluded that a 30-day time frame (typical for assessing prodromal symptoms) would be sufficient to capture the presence of intermittent episodes. In addition, we selected a uniform 'in the past 30 days' time frame rather than a 'past month' time frame to avoid confusion among people who might be thinking about the most recent named month (e.g., September) while answering questions.

In order to define and represent concepts associated with the prodromal period, early psychosis, and psychosis, we included concepts relating to positive symptoms, negative symptoms, and the exclusionary criteria listed in the DSM-5, as well as general symptoms that have been associated with conversion. We then subdivided the items into category 'bins' to assure adequate coverage of related concepts. Concepts included the symptoms listed in DSM-5, such as delusions, hallucinations, disorganized speech, gross disorganization, avolition, and a decrease in functioning; yet, for our purposes, the DSM-5 nomenclature was not sufficiently specific. The term 'delusion' alone, for example, can refer to any number of phenomena: paranoid delusions, persecutory delusions, religious delusions, grandiose delusions, delusions of control, thought insertion, telepathy, thought broadcasting, erotomania, and somatic delusions, to name a few. Drawing from the well-established instruments, we made our 'concepts' as granular as possible. We recognized that overlap in the nomenclature and categories was inevitable and that partial redundancy was, in fact, desirable. In addition, since the DSM-5 criteria for schizophrenia include exclusions relating to schizoaffective disorder, bipolar disorder, and substance abuse, we also included items on depression, anxiety, mania, and substance use for exploratory purposes.

The item pool was iteratively reviewed and rated by a panel of eight experts, including three psychiatrists and three psychologists with combined expertise in the prodromal period and early psychosis, SIPS and SCID (Structured Clinical Interview for the DSM) administration, community mental health, and biostatistics. Other members of the panel included an English professor, a linguist, and two TeleSage, Inc. interns. The panel members reviewed the items for breadth of coverage across the concepts. In addition, each item was rated independently by each panel member on a 3-point scale for clarity and centrality (i.e., 1 = neither clear, nor central to the concept; 2 = clear, but not central to concept OR central to concept, but not clear; and 3 = clear and central to concept). Experts participated in focus groups, where they were asked to describe the benefits and/or problems associated with each item and provide a rationale for each item rating. Experts were also asked to rank similarly worded items in order of their preference. We averaged the results from the expert panel ratings, and considered elaborations provided through the experts' comments and rankings.

Please cite this article as: Brodey, B.B., et al., The Early Psychosis Screener (EPS): Item development and qualitative validation, Schizophr. Res. (2017), https://doi.org/10.1016/j.schres.2017.11.027

2.2. Stage II: cognitive interviewing

All research subjects participated in an IRB-approved informed consent process. Consistent with North American Prodrome Longitudinal Study (NAPLS) practices, participants aged 12–35 were recruited from a Centerstone clinic in Bloomington, Indiana; as well as at the TeleSage site in Chapel Hill, North Carolina (Addington et al., 2012). Participants at the Centerstone site ($n = 26$, average age of 25.5) were drawn from their sample of outpatients. Participants from the TeleSage sample ($n = 16$, average age of 21.1) were recruited from among the SIPS 3,4,5 CHR patients being followed at the NAPLS site at the University of North Carolina, Chapel Hill as part of a separate research study. A total of 42 participants received CIs. Participants ranged in age from 14 to 35 with an average age of 23 years. Of the 42 participants: 67% (28) were female; 69% (29) were white; 14% (6) identified as black; 2% (1) identified as a Hawaiian native; and 5% (2) identified as being ethnically Latino.

CI is a scientific technique that uses 'verbal probes' and 'verbal think alouds' to determine the perceived meaning of survey questions (Willis, 2005). Interviewees are asked to answer 10 written questions at a time. For each question, participants are asked: 'what do you think the individual question means?,' 'what do individual words and phrases mean?,' 'what information do you think the author was actually trying to elicit?', 'what were you thinking as you retrieved from memory the information needed to determine the correct answer to the question?', 'how did the response options fit the question?,' and 'what was your decision process in choosing how to actually answer the question?' (based on superimposed perceptions of social desirability, etc.). By having participants describe all of their thoughts out loud as they work their way through questions, it is possible to identify many of the problems that could affect a patient's response in unintended ways. Using CI to hone questions makes it much more likely, when entering into quantitative testing, that individual items will ultimately have good psychometric characteristics. This was especially important in this study as we knew that we needed to cover a large number of concepts efficiently.

CI was conducted in three rounds with the Centerstone sample and two rounds with the TeleSage sample. (Fewer rounds were conducted with the TeleSage sample given that the Centerstone rounds preceded the TeleSage rounds and led to significant revisions/omissions prior to the start of the TeleSage rounds.) Each round was conducted with unique participants who engaged in an individual interview; no participant was interviewed twice. Each CI took approximately 90 minutes to complete and interviews were audiotaped. The CIs provided ample opportunity for open exploration of items and responses. After each interview, the interviewer summarized her notes from the CI, specifically noting which questions were easier or harder for participants to understand and which questions appeared to represent greater relative severity. A TeleSage staff member transcribed the responses for any items that presented difficulty. The expert panel then re-rated the item based on the feedback from the participants and the interviewer. This process was repeated throughout all five rounds of interviewing.

Interview data was evaluated between the five rounds by the same expert panel to identify items in need of revision or omission from the item pool. After each round of cognitive interviewing, we omitted or revised questions that created difficulties in regards to either interpretation (e.g., were unclear or were interpreted in an unintended manner) or utility (were endorsed with an unexpectedly high base rate). Initially, any item that presented difficulty for at least 20% of participants was deleted or revised. After the second round of interviewing, items that presented no difficulty for all participants were considered complete and removed from the CI process. This afforded more time in subsequent rounds for us to evaluate new and revised items as well as items for which some difficulty was noted. Our goal was that by the end of the final round of interviewing we would be left with a group of items that were well understood by all participants and which would still represent all the concepts.

3. Results

3.1. Part I: item development

The 490 items were divided into 25 general groups. These 'bins' were not intended to be mutually exclusive and were used primarily for convenience in the cognitive interviewing process. For example, De-realization and Erotomania are not similar symptoms; however, Erotomania did not have enough items to warrant an individual bin. Thus, we grouped Erotomania and De-realization items in order to keep the size of the comparison groups consistent. The bins were: (1) Demographics; (2) De-realization; (3) Thought Broadcasting and Telepathy; (4) Ideas of Reference; (5) Hyper-religiosity; (6) Somatization; (7) Persecutory and Paranoid; (8) Grandiosity; (9) Thought Insertion, Deletion, and Guilt; (10) De-realization and Erotomania; (11) Mild Auditory Hallucinations; (12) Prominent Auditory Hallucinations; (13) Visual and Tactile Hallucinations; (14) Affective Flattening; (15) Avolition and ADLs; (16) Concentration and Confusion; (17) Disorganization; (18) Drug Use; (19) Alcohol Use; (20) Depression; (21) Anxiety; (22) Mania; (23) Social Functioning; (24) Work and School Functioning; (25) Social Functioning Relative to Period 12 months prior.

Each domain included multiple concepts. For example, the Hyper-religiosity domain included specific beliefs about (a) being called to do God's work or (b) the devil's work. We also created multiple similar versions of many items in order to ensure that we could identify a version that would work well. For example, as it was particularly difficult to differentiate normal strong religious convictions from delusional beliefs, we wrote and explored many items covering these two concepts (see Table 1 for an example of a Religious Delusion item that addressed these difficulties and was ultimately included). Overall, we developed a much larger item pool than we would ultimately need because we wanted to explore concepts in detail. Panel members performed three rounds of item evaluation before CI began. In each case, items with an average rating of 2.5 or less were eliminated or revised. Items were retained if the panel gave the item an average rating of greater than 2.5 and there was no other item covering the same core concept that had a higher rating. At the end of these three rounds of evaluation, the panel approved 301 items for CI.

3.2. Part II: cognitive interviewing

The first round of cognitive interviewing started with 301 items and took place at Centerstone. We found many items that were interpreted in more than one way and were therefore eliminated (see Table 1 for examples of the kinds of interpretive difficulty we encountered). We retained items about which we had uncertainties to obtain additional feedback. In all, 77 items were omitted, 30 new items were created based on feedback, and 2 items were revised. We started Round 2 at Centerstone with 254 items. Of these items, 34 were omitted, 15 new items were created based on feedback, and 14 items were revised. We started Round 3 at TeleSage with 235 items. Of these items, 37 were omitted, 7 were created, and 12 were revised. We started Round 4 at Centerstone with 205 items: of these 33 were eliminated, and 8 items were slightly revised. On Round 5 at TeleSage, we started with 172 items and eliminated 24 items, which were misunderstood by at least one participant. We ended CI with 148 items, excluding demographics. These 148 items covered the full range of core concepts that we had identified. Each of these 148 items went through at least one full round of CI without generating any confusion. A minimum of 2 cognitive interviews were performed on all of the items. Some items were included in up to 5 rounds of cognitive interviewing. For specific examples of items that were omitted, revised, or retained, see Table 1.

Please cite this article as: Brodey, B.B., et al., The Early Psychosis Screener (EPS): Item development and qualitative validation, Schizophr. Res. (2017), https://doi.org/10.1016/j.schres.2017.11.027

Table 1
Examples of CI results: participant think aloud (TA) and interviewer probing (IP).

Sample omitted items	Reason for omission
I felt the presence of evil around me. (Religious/persecutory delusion)	Responses from participant think aloud (TA) and interviewer probing (IP) indicated that participants interpreted the item as meaning there were 'bad people' (a bad element) around them, which led to a higher base rate of endorsement than was expected.
I thought people might be able to read my mind. (Thought broadcasting/delusions of control)	In response to TA & IP, participants provided examples of parents' knowing what they were thinking and anticipating their next move. This led to a higher base rate of endorsement than was expected.
People said I did not show emotions. (Affective flattening)	Participants stated that people did not say this.
I spent time with friends after work or school. (Social functioning)	TA & IP responses, as well as responses to other social functioning items, indicated that negative endorsement of the item (i.e., Never) often reflected the inability to engage with friends due to other commitments (e.g., do homework, babysit siblings), rather than poor social functioning.
Sample revised items	**Reason for revision**
Original item: I thought people were watching me. Revised item: People were plotting against me. (Paranoid/persecutory delusions)	Participant TA & IP responses indicated a high base rate of endorsement, especially among adolescents. They thought people were watching them all the time at school.
Original Item: I felt anxious. Revised Item: I had anxiety. (Anxiety)	Anxious in the context of "I felt anxious" was defined as excited or eager (e.g., "I was anxious to go to the fair") by participants, particularly those in the South. The noun form, however, did not have the same additional connotation; therefore, the item was revised to use the noun form of anxiety.
Original Item: I thought I might be God's personal messenger on Earth. Revised Item: I am the only person who can do God's work on Earth. (Religious delusions)	The original item produced a high base rate of endorsement among devoutly religious participants. The revised item is distinct from the notion that all people are God's children or messengers.
Original Item: I thought people were planning to hurt me. Revised Item: I thought people were planning to physically hurt me. (Paranoid/persecutory delusions)	Participant interpretations of the original item included emotional harm, which had a high base rate (i.e., hurting feelings). The revised item narrowed the item's focus to physical harm.
Original item: I thought I had special powers. Revised item: I thought I had superhuman powers. (Grandiose delusions)	Participant TA & IP responses indicated a high base rate of endorsement due to personal qualities (e.g., the gift of gab, good listener). The revised item clarifies the item's intent of assessing grandiosity.
Original instructions: Now I'm going to ask you about things you thought you might have seen while you were fully awake and it was light. Revised instructions: Now I'm going to ask you about things you might have seen while you were fully awake and there was enough light to see clearly. (Visual hallucinations)	Participant TA & IP responses indicated high endorsement due to appearance of shadows due to dim light. The revised instructions clarify that visual hallucinations were present when enough light was present to see clearly (i.e., eliminate shadows).
Sample retained items	**Reason for retaining item**
I thought I deserved to be punished. (Persecutory delusion or guilt)	Easily understood in early rounds of CI & central to Persecutory Delusions
My thoughts were being controlled. (Delusions of control)	Easily understood in early rounds of CI & central to Delusions of Control
I felt like someone was touching me, but no one was there. (Tactile hallucinations)	Easily understood in early rounds of CI[a] & central to Somatic Delusions
I had difficulty feeling emotions. (Affective flattening)	Easily understood in early rounds of CI & central to Affective Flattening

[a] This is especially important given the contingent nature of the question.

4. Discussion

We accomplished our goal of using qualitative validation techniques to develop an item bank. In achieving this goal, we encountered three primary difficulties that are summarized here. First, it is difficult to write questions that are simple and which also define an experience or belief in sufficient detail. Second, people have a very broad range of beliefs and experiences. Typically, only constellations of experiences and beliefs represent pathology. Third, most of the words we use to describe experiences and beliefs have developed over hundreds of years and their meanings have evolved and changed during this period. There is no simple taxonomy that clearly defines their unique meanings or commonly understood hierarchies.

Relating to the first difficulty, developing self-report items is an imperfect process. As an example, consider auditory hallucinations. We could ask: 'I heard voices,' but normal individuals hear the voices of other people throughout the day. We could ask: 'I heard voices when I was alone,' but a normal person might be watching television or there might be a loud person in the next room. We could ask: 'I heard voices that nobody else could hear,' but a normal person might simply believe she has very good hearing or, conversely, a person experiencing auditory hallucinations might think that other people hear the voice too. We could ask: 'I heard voices, but I was not sure if they were real,' but a person experiencing auditory hallucinations might feel quite confident in the voices' reality. We also needed to make sure that these experiences have not occurred as the person was waking up, falling asleep, doing drugs, or febrile. These latter qualifiers are probably best dealt with in the instructions as a patient prepares to take the assessment, but a patient also needs to be reminded of the instructions periodically while responding to questions. We could combine specific qualifying statements such as, 'I heard voices that nobody else could hear, when I was alone, and I was not watching the TV.' However, questions with multiple contingencies require a good working memory, an understanding of punctuation, and considerable mental gymnastics on the part of the patient. They also make interpretation of the responses much more complicated. Based on our CI results, we learned that a good question simply asks for a direct report of the patient's perceptions, thoughts, or feelings. In the case of auditory hallucinations, we found that a series of direct questions can be used to ameliorate some of the concerns raised above. Several questions relating to different aspects of an experience can be asked e.g. 'I heard a voice, but I could not tell if it was real,' 'I thought the voice was real,' 'The voice said mean things about me,' and other specific statements. These items were easily understood, and they mediate the potential for multiple interpretations of the initial item.

In terms of the second concern, the range of normal human experience is broad. Many behaviors related to 'core concepts' occur in high rates among the general population, making their use in assessment tools problematic. For example, many people like consulting fortune tellers, some keep 'lucky charms,' and some are sure that aliens regularly visit earth. These ideas alone do not signify that a person is experiencing prodromal or psychotic symptoms. To avoid false positives and to maximize the information associated with each response, individual items need to be written to avoid high base rates. In addition, a screener will ultimately need to use constellations of item endorsements to assess risk.

With regard to the last difficulty, the lack of a standard taxonomy for concepts describing prodromal, early psychosis, and psychosis presented an interesting challenge. We wanted to develop a group of items that

Please cite this article as: Brodey, B.B., et al., The Early Psychosis Screener (EPS): Item development and qualitative validation, Schizophr. Res. (2017), https://doi.org/10.1016/j.schres.2017.11.027

was sufficiently broad, to be confident that we were not missing important concepts that might be useful in identifying CHR individuals. At the same time, we did not want to have a large number of essentially redundant items. We would have liked to be able to use the DSM-5 in this regard, but we found that its nomenclature is too vague to be very useful in detailed classifications. The standard classification systems of the DSM-5 often did not yield unique assignments. For example, some persecutory delusions could equally be described as paranoid, and many paranoid delusions are grandiose. Overall, we had to rely on existing instruments and the judgments of our expert panel members to determine if the item bank we developed was broad enough to cover the full range of perceptions and beliefs. When a clear choice did not exist, we erred on the side of permitting redundancy between similar but distinct concepts, with the knowledge that we could further winnow items in the future. In future work, we hope to develop a short set of items that can be used to assess respondent bias. In addition, we hope to add items to quantify the severity of fully developed symptoms of psychosis.

5. Conclusions

We developed an item bank that we believe can be used to create an early psychosis instrument that has high specificity and good sensitivity. The item bank includes items to which a non-prodromal, non-psychotic person would mostly respond 'never' or 'rarely.' Additionally, we knew that CHR individuals would only positively endorse a subset of the items provided, based on their own unique experiences and beliefs. Using the item development, modification, and selection process described here, we were able to identify 148 items that were well understood and that our expert panel believes cover the breath of concepts associated with the prodromal period and early psychosis. These 148 items should enable us to develop quantitative algorithms to identify CHR individuals and predict those who will ultimately convert.

Conflict of interest

Drs. Inger and Benjamin Brodey are the sole owners of TeleSage, Inc. All other authors declare that they have no conflicts of interest.

Contributors

Dr. Benjamin Brodey wrote the original manuscript, was a member of the expert panel, designed the study, and wrote the protocol. Drs. Jean Addington and Michael First were consultants throughout this project. Drs. Diana Perkins and Scott Woods assisted with the design of the study. Dr. Elaine Walker assisted with the critique of individual items and sent suggestions for inclusion, exclusion, and revision. Dr. Barbara Walsh was a member of the expert panel. Jennifer Nieri was a member of the expert panel, recruited participants, and administered the SIPS and other clinician rating instruments. Dr. M. Brad Nunn was the site PI at Centerstone. Dr. John Putz was the site coordinator for Centerstone. Dr. Inger Brodey assisted with the study design, was a member of the expert panel, and assisted with the editing of the manuscript. All authors contributed to and have approved the final manuscript.

Role of the funding source

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

Acknowledgements

This study was supported by a National Institutes of Health (NIH) grant #MH094023. We thank Miss Kathryn Elkin, who assisted with the preparation and proof-reading of the manuscript.

References

Addington, J., Cadenhead, K.S., Cornblatt, B.A., Mathalon, D.H., McGlashan, T.H., Perkins, D.O., Seidman, L.J., Tsuang, M.T., Walker, E.F., Woods, S.W., Addington, J.A., Cannon, T.D., 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.

Brodey, B.B., Goodman, S.H., Baldasaro, R.E., Brooks-DeWeese, A., Wilson, M.E., Brodey, I.S., Doyle, M.N., 2016. Development of the Perinatal Depression Inventory (PDI)-14 using item response theory: a comparison of the BDI-II, EPDS, PDI, and PHQ-9. Arch. Womens Ment. Health 19:307–316. https://doi.org/10.1007/s00737-015-0553-9.

Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., Ader, D., Fries, J.F., Bruce, B., Rose, M., PROMIS Cooperative Group, 2007. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med. Care 45 (5):S3–S11. https://doi.org/10.1097/01.mlr.0000258615.42478.55.

Comparelli, A., De Carolis, A., Emili, E., Rigucci, S., Falcone, I., Corigliano, V., Curto, M., Trovini, G., Dehning, J., Kotzalidis, G.D., Girardi, P., 2014. Basic symptoms and psychotic symptoms: their relationships in the at risk mental states, first episode and multiepisode schizophrenia. Compr. Psychiatry 55:785–791. https://doi.org/10.1016/j.comppsych.2014.01.006.

DeWalt, D.A., Rothrock, N., Yount, S., Stone, A.A., 2007. Evaluation of item candidates: the PROMIS qualitative item review. Med. Care 45 (5 Suppl 1):S12–S21. https://doi.org/10.1097/01.mlr.0000254567.79743.e2.

Fusar-Poli, P., Cappucciati, M., Rutigliano, G., Lee, T.Y., Beverly, Q., Bonoldi, I., Lelli, J., Kaar, S.J., Gago, E., Rocchettie, M., Patel, R., Bhavsar, V., Tognin, S., Badger, S., Calem, M., Lim, K., Kwon, J.S., Perez, J., McGuire, P., 2016. Towards a standard psychometric diagnostic interview for subjects at ultra high risk of psychosis: CAARMS versus SIPS. Psychiatry J. 2016:1–11. https://doi.org/10.1155/2016/7146341.

Ising, H.K., Veling, W., Loewy, R.L., Rietveld, M.W., Rietdijk, J., Dragt, S., Klaassen, R.M.C., Nieman, D.G., Wunderink, L., Linszen, D.H., van der Gaag, M., 2012. The validity of the 16-item version of the Prodromal Questionnaire (PQ-16) to screen for ultra high risk of developing psychosis in the general help-seeking population. Schizophr. Bull. 38 (6):1288–1296. https://doi.org/10.1093/schbul/sbs068.

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.

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.

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. https://doi.org/10.1109/TPAMI.2005.159.

Reeve, B.B., 2002. An Introduction to Modern Measurement Theory. National Cancer Institute:pp. 1–67 Retrieved from. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.207.3244&rep=rep1&type=pdf.

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., Bearden, C.E., Cadenhead, K.S., Cannon, T.D., Barbara, A.C., Heinssen, R.K., Seidman, L.J., Tarbox, S.I., Tsuang, M.T., Walker, E.F., McGlashan, T.H., Woods, S.W., 2015. Specificity of incident diagnostic outcomes in patients at clinical high risk for psychosis. Schizophr. Bull. 41 (5):1066–10750. https://doi.org/10.1093/schbul/sbv091.

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

Please cite this article as: Brodey, B.B., et al., The Early Psychosis Screener (EPS): Item development and qualitative validation, Schizophr. Res. (2017), https://doi.org/10.1016/j.schres.2017.11.027