Prodromal Assessment With the Structured Interview for Prodromal Syndromes and the Scale of Prodromal Symptoms: Predictive Validity, Interrater Reliability, and Training to Reliability

by Tandy J. Miller, Thomas H. McGlashan, Joanna L. Rosen, Kristen Cadenhead, Joseph Ventura, William McFarlane, Diana O. Perkins, Godfrey D. Pearlson, and Scott W. Woods

Abstract

As the number of studies related to the early identification of and intervention in the schizophrenia prodrome continues to grow, it becomes increasingly critical to develop methods to diagnose this new clinical entity with validity. Furthermore, given the low incidence of patients and the need for multisite collaboration, diagnostic and symptom severity reliability is also crucial. This article provides further data on these psychometric parameters for the prodromal assessment instruments developed by the Prevention through Risk Identification, Management, and Education (PRIME) prodromal research team at Yale University: the Structured Interview for Prodromal Syndromes and the Scale of Prodromal Symptoms. It also presents data suggesting that excellent interrater reliability can be established for diagnosis in a day-and-a-half-long training workshop.

Keywords: Schizophrenia, prodromal, assessment, early indentification.

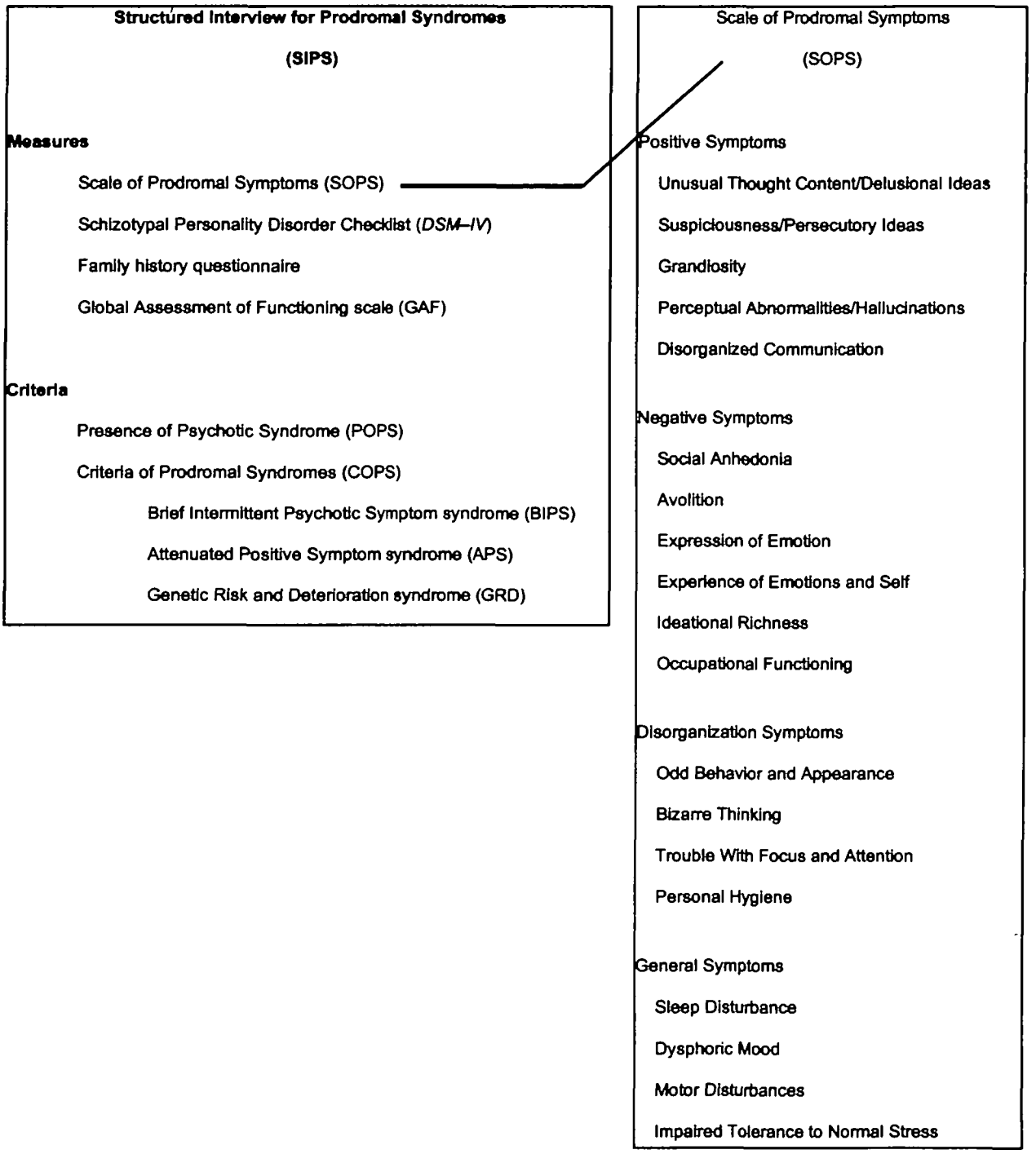
***Schizophrenia Bulletin*, 29(4):703–715, 2003.**

International interest has grown over the past 15 years in the prognostic potential of early identification of and intervention in the prodromal and first episode phases of psychotic illness. As enthusiasm has grown, experts and ethicists have admonished both researchers and clinicians to proceed with caution and with open dialogue concerning the preliminary nature of the evidence base. Most needed is rigorous scientific research, and one of the most critical elements of this research is the capacity to characterize and operationally define the concept of the "prodrome to schizophrenic psychosis" as well as the transition from "prodromal illness" to the onset of psychosis.

Yung et al. (1996, 1998) pioneered this work by outlining three prodromal syndromes that prospectively identified people who were at high risk for developing schizophrenia in the near future. The three syndromes were described as (1) frankly psychotic positive symptoms that appeared too brief and too intermittent to constitute a fully psychotic syndrome, (2) attenuated positive symptoms, and (3) functional decline in the presence of genetic risk. The PRIME prodromal research team at Yale University has developed two instruments to rate and track these phenomena cross-sectionally and over time (see figure 1 for a description of the organization of these instruments and a guide to the acronyms that are used to facilitate communication). These instruments are the Structured Interview for Prodromal Syndromes (SIPS) (McGlashan et al. 2001; Miller et al. 2002; Rosen et al. 2002) and the Scale of Prodromal Symptoms (SOPS) (Miller et al. 1999; McGlashan et al. 2001; Hawkins et al., in press). The SIPS is a structured diagnostic interview used to diagnose the three prodromal syndromes and may be thought of as analogous to the Structured Clinical Interview for *DSM–IV* (SCID) or other structured diagnostic interviews. The SIPS includes the SOPS, the Schizotypal Personality Disorder Checklist (APA 1994), a family history questionnaire (Andreasen et al. 1977), and a well-anchored version of the Global Assessment of Functioning scale (GAF; Hall 1995). The SIPS also includes operational definitions of the three prodromal syndromes (the Criteria of Prodromal Syndromes [COPS]) and an operational definition of psychosis onset (Presence of Psychotic Syndrome [POPS]). As part of the SIPS, the COPS and the POPS are applied to the information from the positive symptoms of the SOPS, the Schizotypal Personality Disorder Checklist, and the family history questionnaire to diagnose a prodromal syndrome or the presence of psychosis. The SOPS is a 19-item scale designed to measure the severity of prodromal symptoms

Send reprint requests to Dr. T. Miller, Yale Psychiatric Research, P.O. Box 208098, New Haven, CT 06520; e-mail: tandy.miller@yale.edu.

Figure 1. Organization of the SIPS and the SOPS and listing of associated acronyms



and changes over time. It may be conceptualized as analogous to the Positive and Negative Syndrome Scale, the Brief Psychiatric Rating Scale, and other established severity rating scales for patients who are fully psychotic. The SOPS contains four subscales for Positive, Negative, Disorganization, and General Symptoms constructs. There are five Positive, six Negative, four Disorganization, and four General Symptoms items. The Negative, Disorganization, and General Symptoms rated on the SOPS are not currently part of making prodromal diagnoses according to

the COPS but are useful in describing the severity of the diagnosis once established. Information regarding the SIPS and the SOPS is available from the authors.

The COPS are listed in table 1. They have been modified from the original criteria developed by Yung et al. in hopes of improving predictive validity by focusing more narrowly on patients at imminent risk. Table 1 also provides a comparison between the COPS and the criteria developed by Yung et al. as currently operationalized in their instrument, the Comprehensive Assessment of At Risk Mental States (CAARMS).

Clinical descriptions of the three prodromal syndromes as defined by the COPS and diagnosed by the SIPS are available in a number of previously published manuscripts (McGlashan et al. 2001; Rosen et al. 2002; Woods and McGlashan, in press). In brief, however, we also include descriptions of the three prodromal syndromes here. The Brief Intermittent Psychotic Symptom syndrome (BIPS) is defined by the experience of frankly psychotic symptoms that do not meet POPS criteria, have reached a psychotic level of intensity only within the past 3 months, and occur at least several minutes per day at a frequency of at least once per month. Thus, clinically, people meeting the BIPS criteria would appear to be experiencing frankly psychotic symptoms of recent onset that are present infrequently and for short periods of time. The Genetic Risk and Deterioration syndrome (GRD) is defined by having a genetic risk, in the form of a first degree relative with any psychotic disorder, or personally meeting the *DSM–IV* criteria for schizotypal personality

Table 1. Comparison of COPS for prodromal patients with CAARMS[1]

Prodromal criteria	CAARMS	COPS
Brief Intermittent Psychotic Symptom syndrome	Severity score in the psychotic range on 1 or more of 3 CAARMS positive items AND When duration more than 1 hr per occasion, frequency less than 3 times per week When duration less than 1 hr per occasion, frequency less than daily AND Episode duration less than 1 wk AND Symptoms present in last yr and for less than 5 yrs	One or more of the 5 SOPS positive items in the psychotic range (rating of 6) AND Symptoms beginning in the past 3 mos AND Symptoms occurring currently at least several minutes per day at least once per mo
Attenuated Positive Symptom syndrome	Severity score in the prodromal range on 1 or more of 3 CAARMS positive items AND When episode at least 1 wk long and duration less than 1 hr per occasion, frequency more than twice per wk OR Frequency once per month and duration less than 1 hr per occasion and more than 2 occasions AND Symptoms present in last yr and for less than 5 yrs	One or more of the 5 SOPS positive items scoring in the prodromal range (rating of 3–5) AND Symptoms beginning within the past yr or increasing 1 or more points within the past yr AND Symptoms occurring at least once per wk for last mo
Genetic Risk and Deterioration syndrome	First degree relative with history of any psychotic disorder OR Schizotypal personality disorder in identified patient or relative AND GAF drop of 30% from premorbid level, sustained for 1 mo	First degree relative with history of any psychotic disorder OR Criteria for schizotypal personality disorder met in patient AND GAF drop of at least 30% over the last mo vs 1 yr ago

Note.—CAARMS = Comprehensive Assessment of At Risk Mental States; COPS = Criteria for Prodromal Syndrome; GAF = Global Assessment of Functioning; SIPS = Structured Interview for Prodromal Syndromes.

[1] Yung et al. criteria from CAARMS version distributed May 2002. COPS criteria from SIPS version 4.0 distributed June 2003.

disorder, as well as having a significant drop in functioning as defined by a GAF drop of 30 percent or more over the past year. This syndrome was included in part to capture individuals who may be experiencing a prodromal phase that is predominantly characterized by negative symptoms, which we expect will be reflected in the significant drop in functioning as measured by the GAF.

The final syndrome, and in the PRIME samples by far the most frequent of the three, is the Attenuated Positive Symptom syndrome (APS). This syndrome is characterized by the development or worsening within the past year of mild or attenuated psychotic symptoms that have not yet reached a psychotic level of intensity and have been present at least once per week in the past month. The characteristics of mild or attenuated positive symptoms are familiar to most clinicians but have generally not been the focus of sustained investigative attention until recent years. These patients report experiencing the precursors to delusions, hallucinations, and thought disorder in the form of unusual thought content, perceptual abnormalities, and disorganized speech. Unusual thought content in the attenuated realm can be of a paranoid, grandiose, or other nature and may range from mild to severe but not to psychotic. Clinical examples of this type of symptom include patient reports of considering the possibility, but clearly not believing, that others might be able to read their minds, that they might be able to read others' minds, or that they might be able to predict or determine the future from dreams.

One of the key determinants of a symptom being considered attenuated and not at a fully psychotic level of intensity is the lack of conviction regarding the externally generated, "real" nature of the symptom as well as the maintenance of insight regarding the sense that the experience is, in fact, a symptom. For example, one high school student who was experiencing suspiciousness reported having the feeling that the entire sophomore class in his school was singling him out and watching him. He also reported realizing that this was not possible as soon as he caught the eye of one of his fellow students. Another young woman reported that even though she lived on the third floor of an apartment building in a city and knew that it was not possible for anyone to see directly into her window, she would sometimes feel that people were watching her and would sometimes not get undressed at night. One young man who reported grandiose unusual thought content reported that he had a "weird" feeling that if his coworkers brushed past him, they would have a better day. He was quick to counter, however, that he knew that this was not possible.

Perceptual abnormalities in the attenuated realm can equally be experienced at a mild to a severe but not at a psychotic level of intensity. Patients experiencing such symptoms can report hearing odd noises, such as banging or clicking or ringing; dogs barking when there is no animal present; or their name being called when no one has called them. More severe but still attenuated symptoms have been described as hearing sounds or voices that seem far away or mumbled. People also report experiencing vague perceptual changes such as seeing colors differently, seeing flashes of light, or seeing geometric shapes. People have also frequently reported seeing shadows out of the corner of their eyes or vague ghostlike figures.

Finally, because thought disorder is a subjective experience that is difficult for the observer to assess, the SIPS measures this experience through disorganized speech. Clinically, we look for people who over time have become circumstantial or tangential in their speech, who are using odd words or unusual phrases, or who otherwise are beginning to have difficulty getting the point across.

As captured by the three syndromes, the prodrome for schizophrenic psychosis is conceived as a period of escalating severity of symptoms or functional decline that lies between the end of the relatively asymptomatic premorbid phase and the beginning of the frankly psychotic phase of schizophrenic psychosis (Woods et al. 2001*b*). The prodrome has some similarity on a conceptual basis to "spectrum" and other schizophrenia-related constructs but is sharply distinguished from them. The distinguishing features primarily relate to course and trajectory of illness. The prodrome construct is like schizotypy and schizotaxia in that symptoms are milder than in frank schizophrenia but differs from them in that symptoms are of relatively recent origin and escalating in severity rather than being stable and enduring. The prodrome construct is similar to the concept of "children at risk" in sharing heightened risk for future progression to schizophrenia but differs in requiring that the state be symptomatic, in not requiring that family history of schizophrenia be present, and in connoting greater imminence of risk. The prodrome construct should also be compared and contrasted with *DSM–IV* conceptualizations of fully psychotic disorders that have not been present long enough to meet criteria for schizophrenia or schizoaffective disorder. These *DSM–IV* concepts are psychotic disorder not otherwise specified (NOS), brief psychotic disorder, and schizophreniform disorder. These *DSM–IV* concepts do not overlap with the APS or GRD prodromal syndromes. *DSM–IV* schizophreniform disorder mostly maps to definitions of full psychosis as operationalized either by the SIPS or by the CAARMS. However, as shown in figure 2, some patients who are late in the course of the BIPS prodromal syndrome as defined by the SIPS could simultaneously meet criteria for early *DSM–IV* schizophreniform disorder. For this overlap to occur, the brief intermittent psychotic symptoms would have to have been present between 1 and

Figure 2. Relationship between duration of fully psychotic symptoms and diagnostic criteria for psychotic disorder and brief psychotic syndromes across 3 diagnostic systems

<table>
<tr><th>Duration</th><th colspan="2">Days</th><th>Wks</th><th>Mos</th></tr>
<tr><td>SIPS</td><td colspan="4">SIPS schizophrenic psychosis
= an average of 4 days per wk for a mo
OR
= 1 day if symptoms seriously disorganizing or dangerous</td></tr>
<tr><td></td><td colspan="2">SIPS BIPS
< an average of 4 days per wk, < 3 mos
not seriously disorganizing or dangerous</td><td colspan="2"></td></tr>
<tr><td>DSM–IV</td><td colspan="2">Brief psychotic disorder
= 1 day but < 1 mo
Psychotic disorder NOS
= 1 day, not yet 1 mo</td><td>Schizophreniform disorder
= 1 mo but < 6 mos</td><td>Schizophrenia
Schizoaffective disorder
> 6 mos, including prodrome</td></tr>
<tr><td>CAARMS</td><td>CAARMS BLIPS
< 1 wk</td><td colspan="3">CAARMS Psychosis
> 1 wk</td></tr>
</table>

Note.—BIPS = Brief Intermittent Psychotic Symptom syndrome; BLIPS = Brief, Limited Intermittent Psychotic Symptom group; CAARMS = Comprehensive Assessment of At Risk Mental States; NOS = not otherwise specified; SIPS = Structured Interview for Prodromal Syndromes.

3 months and also meet *DSM–IV* schizophreniform disorder criteria of being present "a significant portion of the time." Also as shown in figure 2, patients whose fully psychotic experience is of sufficiently short duration to meet *DSM–IV* criteria for psychotic disorder NOS or brief psychotic disorder could potentially meet either BIPS prodromal criteria or full psychosis criteria either using the SIPS or the CAARMS. Whether such patients meet prodromal or psychosis criteria depends on frequency of occurrence or severity for the SIPS and on duration for the CAARMS.

Studies of the prodromal phase to first onset depend on the ability to reliably and validly diagnose patients in this phase with standardized criteria. Progress also depends on the ability to track symptomatic changes over time and to define operationally the point of conversion from "prepsychotic" to psychotic levels of symptom intensity. Because psychosis emerges in a dimensional fashion, the specific "point" of conversion from nonpsychosis to psychosis has never been defined. Any definition is, to some extent, arbitrary. Nevertheless, for valid research to proceed, it is necessary for the field to develop a definition that can be operationalized and applied reliably. The criteria for onset of frank psychosis used by our group are the POPS criteria, part of the SIPS interview (figure 1). The POPS requires that one or more of the positive items from the SOPS be scored at a psychotic level of intensity and also describes psychotic symptom frequency and duration criteria. Figure 2 presents the frequency and duration criteria from the POPS along with a comparison to the CAARMS onset of psychosis criteria used by Yung et al.

The PRIME Research Clinic at Yale University has focused on the early identification of and intervention in psychotic illness. Patients are referred to the PRIME Clinic as the result of a variety of community education efforts targeting primarily the psychiatric, medical, and educational systems. Patients who telephone the PRIME Clinic are invited for interview if they meet the following telephone-screening criteria: (1) between the ages of 12 and 45, (2) reporting one or more symptoms that could be prodromal, and (3) not reporting an established diagnosis of psychotic disorder or other Axis I or Axis II disorder that clearly accounts for the possibly "prodromal" symptoms.

The PRIME Clinic has been examining the predictive validity of the SIPS through an ongoing followup study. Initial predictive validity results demonstrated that of 13 patients diagnosed with a prodromal syndrome at baseline, 6 (46%) had developed schizophrenic psychosis by 6

months and 7 (54%) by 12 months (Miller et al. 2002). In addition, none of the initially nonprodromal patients developed schizophrenic psychosis by 12 months. Further updated results from this ongoing study are presented in this article. Five new subjects have been added, and the followup interval for most of the cohort has been extended to 2 years.

Initial interrater reliability data on the SIPS diagnosis of prodromal syndrome have also been reported (Miller 2002). In our reliability study, 7 subjects were diagnosed prodromal and 11 were diagnosed as nonprodromal according to the SIPS. One to three raters per subject in addition to the interviewer made SIPS diagnoses independently from those of the interviewer and other raters by viewing videotapes of the interviews. The agreement among raters was 93 percent for the judgment of whether the subjects were prodromal or nonprodromal (kappa = 0.81, 95% confidence interval = 0.55–0.93). This study demonstrates that the interrater reliability was excellent for diagnosis. We have completed a new study examining interrater reliability among raters for the individual items on the SOPS. Results will be described in this article.

Investigators in the field of prodromal research must learn how to identify what, in essence, is a new clinical entity. This, in turn, requires rater-training workshops built around the structured diagnostic interview. We know that diagnostic measures and operational criteria for diagnostic classification improve the reliability of diagnosis (Feighner et al. 1972; Luria and Berry 1979; Andreasen et al. 1982; Williams et al. 1992; Ventura et al. 1998). We also know that standardized rater-training programs have been shown to teach clinicians and researchers with varying levels of clinical experience to use these structured interviews reliably (Flemenbaum and Zimmerman 1973; Luria and Berry 1980; Ventura et al. 1993; Gutkind et al. 2001). The PRIME Clinic has created such a training program and conducted six workshops to date. We will describe this training program and the participants' levels of reliability before and after the workshop.

In this article, we will first update the predictive validity study. Second, we will describe the training program that we developed for SIPS interviewers and its results in improving diagnostic reliability conducted at six sites. Finally, we will report on the reliability study of the individual SOPS items.

Methods

Predictive Validity Study

Subjects. Patients were drawn from a total of 123 consecutive symptomatic, treatment-seeking individuals who were referred to the PRIME Clinic for a suspected prodromal syndrome, gave written informed consent, and were given the SIPS from January 23, 1998, through September 1, 2002. Of these patients, 55 were ineligible for the predictive validity study. Of these 55 ineligible patients, 41 had entered clinical trials (all 41 had been diagnosed as prodromal at baseline), 12 met POPS criteria for psychosis, and 2 were missing baseline data on prodromal status. The remaining 68 were eligible for the predictive validity study; however, thus far only 34 (50%) have participated.

The mean age of the 34 followup subjects was 17.9 years (SD = 5.8), and 23 (68%) were male. Of these 34, 14 met the criteria for the prodromal syndrome at baseline, and 20 did not meet the criteria for either psychosis or prodromal syndrome. Of the 34 nonparticipants in the validity study, 12 are waiting to be scheduled, 12 could not be located, 9 refused to participate usually because they felt the time commitment to be burdensome, and 1 was deceased. The mean age for these 34 nonparticipants was 21.8 years (SD = 12.0); 22 (65%) were male, and 14 had prodromal syndromes.

Procedures. The SIPS was part of a face-to-face interview that was conducted at baseline and repeated over evaluation periods of 6, 12, 18, and 24 months. In four cases, followup interviews were conducted over the telephone. Patients who were diagnosed at baseline as prodromal were considered still prodromal unless they had developed psychosis or had remitted (defined as the absence of any positive symptom item in the SOPS with a score in the prodromal range). Interviewers were all "certified" SIPS raters. To be certified, each interviewer must have previously participated in sessions with at least five symptomatic prodromal patients where one of the SIPS developers was the primary rater and must be judged by one of the SIPS developers as competent to administer the SIPS independently (Miller et al. 2002).

Rater-Training Study

Procedures. A standard one-and-a-half-day training program to teach participants to use the SIPS was developed at the PRIME Clinic by the developers of the measures. The program consists of five phases. During the orientation phase, the participants are provided with looseleaf notebooks that contain copies of the measures, articles that describe the measures, and copies of the slides from an introductory lecture and an advanced lecture that addresses specific rating issues. The notebooks also include the transcripts from four different videotapes, each accompanied by its own copy of the SIPS.

On the morning of the first day, the trainees are given an introductory lecture that describes the measures in detail. This portion of the training is intended to familiarize the participants with the measures enough that they are oriented to the general goals, format, and characteris-

tics of the instrument. In addition, the lecture involves a discussion that includes clinical examples of the three types of prodromal syndromes defined by the SIPS that is similar to the one presented in this article. Finally, relevant preliminary psychometric studies are briefly presented and reviewed.

The trainees are then shown and asked to rate two different videotapes: one of a patient with prodromal symptoms and one of a patient who does not meet prodromal criteria according to the SIPS. The trainees are blind to the diagnosis of each patient as established by the trainer. The videotapes are discussed at length in an item-by-item format during the afternoon of the first day.

On the morning of the second day, the trainees are given a lecture that highlights specific rating issues and reviews and summarizes the issues discussed the previous day. Some of the most salient points that are discussed in this lecture are presented here. First, the concept of "delusional conviction," which is critical to determining the difference between a severe but not psychotic rating versus a severe and psychotic rating, is explained. "Delusional conviction" is considered to be the experience of believing that an experience is either "real" or external to oneself without any doubt, at least momentarily. Patients who experience conviction report that they believe that, for example, the voice that they are hearing is not being generated by their own brain but rather exists independently and externally. Individuals who are not experiencing delusional conviction offer other explanations for their experience. One common comment that is given to explain a symptom is that, because they believe that the experience cannot be real, they claim their mind must be "playing tricks" on them. The raters are instructed, when using the interview, to query the patients as to how they account for the experience, whether they believe that it is real or "just in their head." In the case of voices, to establish the experience of an hallucination as differentiated from a thought, interviewers are instructed to ask such questions as whether the patient can hear the voice/sound with his or her ears, whether someone else may be able to hear it, and whether the voice/sound is heard "out loud" as clearly as the interviewer's voice is experienced.

The above discussion helps to differentiate between a severe but not psychotic symptom and a severe and psychotic symptom. We also address the distinction between a moderately severe symptom and a severe but not psychotic symptom. We point out that the severe but not psychotic rating is reserved for patients who are right on the edge of believing that their experience is real or external but who are able to acknowledge that they have some doubt or that doubt can be induced by contrary evidence and others' opinions.

One factor that has proven to be slightly confusing to raters is how to rate a persecutory versus nonpersecutory idea of reference. The raters are simply reminded that persecutory ideas of reference are rated under Suspiciousness (scale item P.2) and nonpersecutory ideas of reference are rated under Unusual Thought Content (scale item P.1). The raters are reminded of some of the other information that is provided in the SIPS instructions but is sometimes overlooked as well as some general issues related to rating scales. They are reminded that the interview inquires about the experience of the patient over the lifetime but that the ratings on the SOPS refer to the patient's experience over only the past month. Also, they are reminded that the SIPS and SOPS are designed to be phenomenological measures. In other words, the rater is not intended to change any rating based on what he or she believes the cause of the patient's experience may be. If, however, each symptom otherwise qualifying for a prodromal diagnosis is better explained by another *DSM–IV* diagnosis, the patient may not be diagnosed with a prodromal syndrome. For this reason, we now recommend that the SIPS be used in conjunction with another, more general diagnostic measure, such as the SCID. Along similar lines, the raters are told that ratings are not intended to be made based solely on information provided by a collateral source but rather should be based on interviewer observation and/or patient report. When the collateral and patient sources of information disagree, we suggest that the collateral informant be present in the interview to confront the patient with the discordant information.

Raters are also reminded that when in doubt over a particular rating, they should rate to the extreme, keeping in mind that the moderate to severe but not psychotic range is considered to "put" a symptom into the attenuated psychotic range of intensity. Finally, the anchors, provided to guide raters to choose a particular level of intensity for any given symptom and to maximize interrater reliability, cannot enumerate every possible symptomatic experience, especially because of the heterogeneous nature of the experiences of patients in the prodromal phase of psychosis. Thus, raters are encouraged, when in doubt, to revert to the general descriptors of different scale points such as mild and moderate.

The events of the first day and the morning of the second day constitute the training portion of the rater-training workshops that are reported in this article. At the end of the training workshop, the trainees are then asked to view and rate two new videotapes; again, one videotaped patient has prodromal symptoms and the other does not meet criteria for a prodromal syndrome according to the SIPS. This set of second ratings is intended to measure the effect of the rater-training workshop. Again, the trainees are blind to the established diagnosis. The

remainder of the second day is spent answering any further questions that arise and any of the items that participants are interested in discussing further, even though it is not possible to measure the effect of these further discussions.

Trainer and videotapes. Tandy J. Miller, a licensed clinical psychologist and one of the original developers of the SIPS and the SOPS, provided all of the training. All of the patients who attend interviews at the PRIME Clinic are invited to have their baseline interviews videotaped. If they assent, they are asked to sign a consent form authorizing the principal investor to use the videotapes for purposes relevant to this research, including training. When the patient is a minor, the parent or guardian is asked to give consent and the minor to give assent. The same set of four videotapes, in the same order, was used in each of the rater-training workshops. The specific videotapes were chosen based solely on an attempt to maximize the sound and recording quality of the taped interviews. The tapes are judged by other certified raters to be of a consistent level of difficulty in terms of diagnostic determination.

Participants. Rater-training workshops were conducted six different times at five sites with 35 raters. The sites were the Maine Medical Center, Portland, ME (MMC: 7 raters); the University of California San Diego, San Diego, CA (UCSD: 5 raters); the University of California Los Angeles, Los Angeles, CA (UCLA: 7 raters); the Institute of Living, Hartford, CT (IOL: 8 raters), the Maine Medical Center, Portland, ME (Maine: 2 raters not trained with the first group); and the University of North Carolina, Chapel Hill, NC (UNC: 6 raters). The participants in the rater-training workshops were all native, English-speaking mental health professionals. The group included licensed psychologists, psychiatrists, master's level health care professionals, nurses, and research assistants.

"Gold-standard" ratings and analysis. The four training videos that are accompanied by "gold-standard" ratings, as determined by Dr. Miller, are used for computing pretraining videos 1 and 2 and posttraining videos 3 and 4 interrater reliability and in providing the trainees with feedback. We use the kappa statistic to compare trainee agreement with the gold-standard diagnosis on presence or absence of a prodromal syndrome (Cohen 1960). Definitions of clinical or practical significance for kappa and r_{icc} have been proposed (<0.40, poor; 0.40–0.59, fair; 0.60–0.74, good; <0.75, excellent) Cicchetti and Sparrow 1981).

SOPS Interrater Reliability Study

Subjects. SOPS interrater reliability was examined in 14 consecutive baseline interviews of subjects who were evaluated as being prodromal for psychosis when two or more raters were present between March 3, 2000, and March 21, 2001. These subjects ranged in age from 12 to 26, with a mean age of 16.5. There were seven females and seven males. Six were Caucasian, three were of African descent, four were of Hispanic descent, and one was of Asian descent.

Raters. The raters for this reliability study were one licensed, clinical psychologist (T.J.M.) and two psychology postdoctoral fellows (including J.L.R.). The interviewer was one of the raters in each case. The interviewer asked about the subject's symptoms as per the SOPS procedures. The interviewer was one of the SOPS developers (T.J.M.) or a certified interviewer (J.R., L.S.). All raters were trained in the use of the SOPS through an apprenticeship model, as described above.

Procedures. Ratings for this reliability study were made in person by two to three raters who were present during the session. The number of raters varied from two to three, depending on the number of people available to be present at a particular session. Ratings for this reliability study were made and recorded independently and not discussed among raters.

Analysis. Intraclass correlations were generated using a computer program that permits calculation of r_{icc} when the number of raters is unequal across subjects (Cicchetti and Showalter 1988).

Results

Predictive Validity. In the validity study, 13 of the 14 initially prodromal patients met the prodromal syndrome defined by attenuated positive symptoms at baseline, and 1 met the prodromal syndrome defined by brief intermittent psychotic symptoms. No patient met GRD criteria in this sample. Figure 3 shows that 6 of the 14 (positive predictive value [PPV] = 43%) developed schizophrenic psychosis by 6 months and that the PPV was 50 percent at 12 months. Sensitivity and specificity were 100 percent and 71 percent, respectively, at 6 months and 100 percent and 74 percent at 12 months. At 18 months the PPV was 62 percent (8 of 13) and at 24 months 67 percent (8 of 12). Sensitivity and specificity were 100 percent and 74 percent, respectively, at 18 months and 100 percent and 73 percent at 24 months. Of the 7 initially prodromal patients at baseline who did not convert by 12 months, 5 of these remained prodromal, 1 remitted, and 1 developed a new *DSM–IV* disorder (major depression). The difference in conversion rates between initially prodromal and initially nonprodromal subjects was highly significant at each time point using Fisher's exact test ($p < 0.005$ at 6 and 24 months, $p < 0.001$ at 12 and 18 months). Of the 20 initially nonprodromal patients, none developed schizo-

Figure 3. Conversion rates to schizophrenic psychosis and diagnostic efficiency measures at 6, 12, 18, and 24 mos among 34 subjects who underwent baseline SIPS interview

SIPS interview

	6-mo converters		6-mo nonconverters		Total	
SIPS+	6	(18%)	8	(24%)	14	(41%)
SIPS–	0	(0%)	20	(59%)	20	(59%)
Total	6	(18%)	28	(82%)	34	(100%)

Incidence	18%
Sensitivity	100%
Specificity	71%
Positive predictive value	43%
Prediction ratio	Inf.

	12-mo converters		12-mo nonconverters		Total	
SIPS+	7	(21%)	7	(21%)	14	(41%)
SIPS–	0	(0%)	20	(59%)	20	(59%)
Total	7	(21%)	27	(79%)	34	(100%)

Incidence	21%
Sensitivity	100%
Specificity	74%
Positive predictive value	50%
Prediction ratio	Inf.

	18-mo converters		18-mo nonconverters		Total	
SIPS+	8	(30%)	5	(19%)	13	(48%)
SIPS–	0	(0%)	14	(52%)	14	(52%)
Total	8	(30%)	19	(70%)	27	(100%)

Incidence	30%
Sensitivity	100%
Specificity	74%
Positive predictive value	62%
Prediction ratio	Inf.

	24-mo converters		24-mo nonconverters		Total	
SIPS+	8	(35%)	4	(17%)	12	(52%)
SIPS–	0	(0%)	11	(48%)	11	(48%)
Total	8	(35%)	15	(65%)	23	(100%)

Incidence	35%
Sensitivity	100%
Specificity	73%
Positive predictive value	67%
Prediction ratio	Inf.

Note.—Inf. = infinite; SIPS = Structured Interview for Prodromal Syndromes.

Table 2. Agreement between trainees and trainer on SIPS diagnosis before and after a 2-day rater-training workshop[1]

Measure	Session	Site						
		MMC	UCSD	UCLA	IOL	Maine	UNC	All sites
Agreement	Before	0.571	0.400	0.429	−0.250	0.500	0.500	0.314
With trainer	After	0.714	0.800	0.857	1.00	1.00	0.833	0.857

Note.—IOL = Institute of Living; Maine = Maine Medical Center (2 raters in second group); MMC = Maine Medical Center (7 raters in first group); UCLA = University of California, Los Angeles; UCSD = University of California, San Diego; UNC = University of North Carolina, Chapel Hill.

[1] Agreement expressed as kappa values (< 0.40, poor; 0.40–0.59, fair; 0.60–0.74, good; ≥ 0.75, excellent).

phrenic psychosis. Although none of these 20 initially nonprodromal patients ever met psychosis criteria, two of them newly met prodromal criteria by 12 months and a third by 18 months. Two of these three subjects then entered a clinical trial, so data on further medication-free followup could not be obtained. The third has not been followed beyond 12 months so far. Additional evidence that patients who did not develop schizophrenic psychosis were correctly diagnosed is that only two (378, 396) received antipsychotic medication during followup so far, and in both of these cases the indication was not psychosis.

Rater Training. Table 2 shows diagnostic agreement before and after training at each site. The data demonstrate that agreement with the "gold standard" before training was poor among all 35 trainees as well as poor or fair at each individual site. Posttraining data, in contrast, reveal that agreement among all 35 trainees at the six different sites was in the excellent range. In addition, per site posttraining data demonstrate that agreement among raters with the "gold standard" in five out of the six sites was in the excellent range and near the excellent range in the final site.

SOPS Interrater Reliability. Reliability analyses on the SOPS severity rating scale revealed that the r_{icc} value was 0.95 for the total score and above 0.75 for all four subscales (Positive, Negative, Disorganized, and General Symptoms subscales), thus being in the excellent range for the total score and all subscales. For individual items, agreement was in the excellent range for 17 out of 19 and near the excellent range for the other 2 (0.70 and 0.72).

Discussion

The relationship between the initial SIPS diagnosis and the 24-month followup data supports the predictive validity of the COPS when used in association with the SIPS. The overall rate of transition to schizophrenic psychosis at 12 months observed in our small sample (7 of 14, 50%) is similar to that observed in a larger sample (20 of 49, 40.8%) (Yung et al. 2003). Because the COPS used in our PRIME Clinic were based on the Yung et al. criteria, the similarity in predictive validity results suggests that prodromal diagnostic criteria may be applied consistently across sites.

One methodological consideration in comparing the current findings with the Yung et al. (2003) findings relates to differing definitions of the exit point to psychosis in these two longitudinal studies of the prodrome. The differences relate to differing symptom measures to determine whether full psychosis criteria are met, as well as to differing duration criteria of fully psychotic symptoms. For the Yung et al. criteria, the symptom measures originally were based on the BPRS and the Comprehensive Assessment of Symptoms and History. The symptom measure currently used is the CAARMS. In our work with the POPS criteria, the symptom measure has been the SOPS. The duration criteria differences are shown in figure 2. Inspection of the POPS criteria and the Yung et al. criteria suggest that they are likely to be strongly overlapping, but empirical studies have not been conducted to determine how often the same patients would meet both sets of criteria for onset of psychosis at the same time.

Another consideration in comparing the current results with those from Yung et al. (2003) relates to diagnostic classification of the nonconverters. In the comparison sample, 12 of 29 nonconverters (24.5% of the total sample of 49) were reported to have no diagnosis at 12-month followup, and 15 of 29 (30.6%) were reported to have a nonpsychotic diagnosis. In our study, diagnostic classification at followup provided for the possibility that patients could be classified as still meeting prodromal criteria, and 5 of 7 nonconverters met such criteria. These patients are believed to be still at risk for conversion, and 1 of 5 did convert by 18 months. If these cases are considered as remaining at risk and thus not yet false positives, the proportion of "true false positives" at 12 months is only 2 of 14 (14%).

Other differences in methodological procedures are evident in table 1. Although the two sets of prodromal criteria in table 1 are generally similar, the COPS in general are somewhat more restrictive. The CAARMS permits

patients with relatively stable symptoms over the past 5 years to meet criteria, while for APS the COPS require patients to be getting worse in the past year. For BIPS, the COPS require that symptoms have reached the psychotic range in the past 3 months. For GRD, the CAARMS permits patients with a family history of schizotypal personality disorder to meet criteria, while the COPS do not. These differences may explain why BIPS and GRD have been observed less frequently when using the COPS than when using the Yung et al. criteria (Miller et al. 2003).

In addition to followup data on patients initially diagnosed as prodromal, we report followup data on patients who were sufficiently symptomatic to be invited to a SIPS interview but who did not meet prodromal criteria. Other groups have not yet reported similar data. The findings suggest that a diagnosis of "not prodromal" on SIPS interview is strongly predictive of not converting to psychosis by 12 months (20 of 20, 100%). Interestingly, three such patients did develop prodromal syndromes over time, however. These findings suggest that patients initially diagnosed as not prodromal should be urged to return for a followup evaluation if symptoms worsen.

An important limitation of the present predictive validity data, in addition to the small sample sizes, is the relatively low participation rate (50%) that has occurred so far. If the sample that has not yet participated is discovered to show a higher or lower conversion rate for the initially prodromal subjects or a lower conversion rate for the initially nonprodromal subjects, then the sensitivity, specificity, and PPV values shown in figure 3 will be altered. Another limitation is that followup diagnostic assessors were not blind to the results of the initial assessment. From the point of view of providing a stringent test of the validity of the initial assessment, blind followup ratings would have been preferable. On the other hand, the procedure we followed made it possible for the followup interviewer to review the results of the initial assessment and may thus have been preferable from the viewpoint of exploring the course of illness in these patients by increasing sensitivity of detection of psychopathology.

The key to achieving consistent predictive validity is to maximize diagnostic interrater reliability across raters and sites. The present significant results from the training workshop suggest that it is possible to achieve excellent diagnostic interrater reliability with a minimum of training. These results demonstrate that diagnostic agreement across sites is possible when using the SIPS, suggesting that inclusion criteria have the potential to be reliably applied in multisite studies using the SIPS.

Before training, the raters were preliminarily familiar with the instrument and had been oriented to it. The before-and-after training comparison thus represents a fair test of the need for training on the measure beyond a basic orientation program. The poor reliability observed before training suggests that training is needed before raters should use the SIPS to make prodromal diagnoses. This may not be true for every rater and may depend on each rater's level of clinical experience, familiarity with other structured interviews, and familiarity and experience with prodromal patients.

Caveats about the training data are important to emphasize. The training to reliability data presented here are preliminary and are based on ratings of only two videotapes per rater. Further reliability studies would be of value. The training program offered at the training sites evaluated the reliability of trainees to make a diagnostic rating but not their ability to elicit relevant information in an actual interview setting. Assessment of interview reliability rather than rating reliability would require subjects to undergo two or more independent interviews with different interviewers within a short span of time. Although this procedure would impose a substantial burden on the subjects, a study investigating interviewer reliability for the prodromal diagnosis should probably be conducted. Lastly, our training data are limited in that they include only native English-speaking participants. The SIPS and the SOPS have been translated into 14 different languages for use in studies around the world, necessitating the creation of workshops for participants who do not have English as a first language.

Finally, the results from the SOPS interrater reliability study demonstrate that certified raters are able to achieve excellent interrater reliability on the total score, excellent reliability on all four SOPS subscales, and excellent or near-excellent reliability on all of the 19 individual items. These data suggest that the SOPS may be useful as a measure to describe severity of illness among prodromal patients, both at baseline and over time in studies with longitudinal designs. It is useful to note, however, that the training workshop described earlier has been geared toward agreement on prodromal diagnosis for the purposes of enrollment into studies, and not on agreement about the severity of patient's prodromal symptoms. Although we have demonstrated that the interrater reliability within our PRIME site in New Haven is excellent for that aim, it is likely that additional training beyond a day-and-a-half will be necessary to establish reliability for severity of illness.

The followup studies reviewed and the data from our clinic represent converging lines of evidence suggesting that the prodromal diagnosis is a viable concept that can be operationally defined with validity from a prospective viewpoint. Being able to identify the prodrome with reliability and validity strongly supports studies into the natural history, pathophysiology, and possible treatment of this clinical, at-risk syndrome. On a cautionary note, however, it is important to emphasize that the reliability and

validity found in these studies are in part a function of the enriched base rate of conversion in the PRIME Clinic samples. In our predictive validity sample the incidence of new cases of schizophrenic psychosis without regard to the SIPS results was 21 percent per year in the first year, approximately 200 times the annual population incidence rate for this age range (Woods et al. 2001*a*). It is important to emphasize that the results we have observed may not generalize to samples with less enriched base rates.

The imperfect predictive validity, on the other hand, signals the need to find additional markers of risk to minimize the number of false positive identifications. There may well be multiple psychopathological pathways into the first psychotic episode, and the need to map additional pathways needs to be borne in mind, as does variability in content and sequence of symptoms. We are still in a very early stage of measuring this phase of illness from a psychopathological point of view. Larger samples are needed to explore alternative diagnostic cutoff values on positive prodromal symptoms and to explore the possible diagnostic value of other symptomatology, including negative symptoms and self-experienced cognitive decline. In addition to clinically detected symptoms, many other factors may signal dysequilibrium as a harbinger of decline. These factors should be included in future studies to augment information obtained from clinical interviews. Methods that could be able to detect such other factors include stress evaluations, neuropsychological testing, neurological soft signs or minor physical anomalies, structural and functional imaging, electrophysiology, and genotyping. These studies optimally should be conducted before standards to care evolve so that medication-free longitudinal followup is ethically acceptable.

In addition to the SIPS and SOPS, whose psychometric properties are described in this article, other assessment instruments are available and in use for the prodrome. These include the CAARMS (Yung et al. 2003), which has been discussed earlier, and the Bonn Scale for the Assessment of Basic Symptoms (Klosterkötter et al. 2001). Future research on the prodrome, therefore, should also include studies that apply more than one instrument to the same sample. In this way, data concerning overlap and concordant validity can be generated. These studies can enhance the generalizability of the results obtained with any single measure.

References

American Psychiatric Association. *DSM–IV: Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: APA, 1994.

Andreasen, N.C.; Endicott, J.; Spitzer, R.L.; and Winokur, G. The family history method using diagnostic criteria: Reliability and validity. *Archives of General Psychiatry*, 34:1229–1235, 1977.

Andreasen, N.C.; McDonald-Scott, P.; Grove, W.M.; Keller, M.B.; Shapiro, R.W.; and Hirschfeld, R.M.A. Assessment of reliability in multicenter collaborative research with a videotape approach. *American Journal of Psychiatry*, 139:876–882, 1982.

Cicchetti, D.V., and Showalter, D. A computer program for determining the reliability of dimensionally scaled data when the numbers and specific sets of examiners may vary at each assessment. *Educational and Psychological Measurement*, 48(3):717–720, 1988.

Cicchetti, D.V., and Sparrow, S.A. Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86(2):127–137, 1981.

Cohen, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46, 1960.

Feighner, J.P.; Robins, E.; Guze, S.B.; Woodruff, R.A.; Winokur, G.; and Munoz, R. Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, 26:57–63, 1972.

Flemenbaum, A., and Zimmerman, R.L. Inter- and intrarater reliability of the Brief Psychiatric Rating Scale. *Psychological Reports*, 36:783–792, 1973.

Gutkind, D.; Ventura, J.; Barr, C.; Shaner, A.; Green, M.; and Mintz, J. Factors affecting reliability and confidence of *DSM–III–R* psychosis-related diagnosis. *Psychiatry Research*, 101:269–275, 2001.

Hall, R. Global Assessment of Functioning: A modified scale. *Psychosomatics*, 36:267–275, 1995.

Hawkins, K.A.; Quinlan, D.; Miller, T.J.; Woods, S.W.; Zipursky, R.B.; Perkins, D.O.; Addington, J.; and McGlashan, T.H. Factorial structure of the scale of prodromal symptoms. *Schizophrenia Research*, in press.

Klosterkötter, J.; Hellmich, M.; Steinmeyer, E.M.; and Schultze-Lutter, F. Diagnosing schizophrenia in the initial prodromal phase. *Archives of General Psychiatry*, 58:158–164, 2001.

Luria, R.E., and Berry, R. Reliability and descriptive validity of PSE syndromes. *Archives of General Psychiatry*, 36:1187–1195, 1979.

Luria, R.E., and Berry, R. Teaching the Present State Examination in America. *American Journal of Psychiatry*, 137:26–31, 1980.

McGlashan, T.H.; Miller, T.J.; Woods, S.W.; Hoffman, R.E.; and Davidson, L. A scale for the assessment of prodromal symptoms and states. In: Miller, T.J.; Mednick,

S.A.; McGlashan, T.H.; Liberger, J.; and Johannessen, J.O., eds. *Early Intervention in Psychotic Disorders*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2001. pp. 135–149.

Miller, T.J.; McGlashan, T.H.; Rosen, J.L.; Somjee, L.; Markovich, P.J.; Stein, K.; and Woods, S.W. Prospective diagnosis of the prodrome for schizophrenia: Preliminary evidence of interrater reliability and predictive validity using operational criteria and a structured interview. *American Journal of Psychiatry*, 159:863–865, 2002.

Miller, T.J.; McGlashan, T.H.; Woods, S.W.; Stein, K.; Driesen, N.; Corcoran, C.M.; Hoffman, R.; and Davidson, L. Symptom assessment in schizophrenic prodromal states. Psychiatric Quarterly, 70:273–287, 1999.

Miller, T.J.; Zipursky, R.B.; Perkins, D.; Addington, J.; Woods, S.W.; Hawkins, K.A.; Hoffman, R.; Preda, A.; Epstein, I.; Addington, D.; Lindborg, S.; Marquez, E.; Tohen, M.; Breier, A.; and McGlashan, T.H. The PRIME North America randomized double-blind clinical trial of olanzapine vs placebo in patients at risk of being prodromally symptomatic for psychosis: II. Baseline characteristics of the "prodromal" sample. *Schizophrenia Research*, 61:19–30, 2003.

Rosen, J.L.; Woods, S.W.; Miller, T.J.; and McGlashan, T.H. Prospective observations of emerging psychosis. *Journal of Nervous and Mental Disorders*, 190:133–141, 2002.

Ventura, J.; Green, M.F.; Shaner, A.; and Liberman, R.P. Training and quality assurance with the Brief Psychiatric Rating Scale: 'The Drift Busters.' *International Journal of Methods in Psychiatric Research*, 3:221–244, 1993.

Ventura, J.; Liberman, R.P.; Green, M.F.; Shaner, A.; and Mintz, J. Training and quality assurance with the Structured Clinical Interview for *DSM–IV* (SCID–I/P). *Psychiatry Research*, 79:163–173, 1998.

Williams, J.B.W.; Gibbon, M.; First, M.B.; Spitzer, R.L.; Davies, M.; Borus, J.; Howes, M.J.; Kane, J.; Pope, H.G.; Rounsaville, B.; and Wittchen, H.-U. The Structured Clinical Interview for *DSM–III–R* (SCID): II. Multisite test-retest reliability. *Archives of General Psychiatry*, 49:630–636, 1992.

Woods, S.W., and McGlashan, T.H. Special issues in psychosis: Early detection and early intervention. In: Sadock, B.J., and Sadock, V.A., eds. *Comprehensive Textbook of Psychiatry*. 8th ed. Baltimore, MD: Lippincott, Williams & Wilkins, in press.

Woods, S.W.; Miller, T.J.; Davidson, L.; Hawkins, K.A.; Sernyak, M.J.; and McGlashan, T.H. Estimated yield of early detection of prodromal or first episode patients by screening first degree relatives of schizophrenic patients. *Schizophrenia Research*, 52:21–27, 2001*a*.

Woods, S.W.; Miller, T.J.; and McGlashan, T.H. The prodromal patient: Both symptomatic and at risk. *CNS Spectrums*, 6(3):223–232, 2001*b*.

Yung, A.R.; McGorry, P.D.; McFarlane, C.A.; Jackson, H.J.; Patton, G.C.; and Rakkar, A. Monitoring and care of young people at incipient risk of psychosis. *Schizophrenia Bulletin*, 22(2):283–303, 1996.

Yung, A.R.; Phillips, L.J.; McGorry, P.D.; McFarlane, C.A.; Francey, S.; Harrigan, S.; Pattaon, G.C.; and Jackson, H.J. Prediction of psychosis. A step toward indicated prevention of schizophrenia. *British Journal of Psychiatry*, 172(Suppl):14–20, 1998.

Yung, A.R.; Phillips, L.J.; Yuen, H.P.; Francey, S.M.; McFarlane, C.A.; Hallgren, M.; and McGorry, P.D. Psychosis prediction: 12-month follow up of a high-risk ("prodromal") group. *Schizophrenia Research*, 60:21–32, 2003.

Acknowledgments

We would like to acknowledge the assistance and general support of Dr. William Cook, Dr. Rogelio Apiquian, Karen Andersen, and Ellen Rothman. This work was funded by grants from the National Institute of Mental Health #MH60720 (PI: K.S. Cadenhead), #MH60504-04 (PI: G. Pearlson), #MH43775-09 (PI: G. Pearlson), #MH 01905 (PI: D. Perkins), Local Initiatives Funding Partners of Robert Wood Johnson Foundation (PI: W.R. McFarlane, M.D.), The Center for Mental Health Services (PI: W.R. McFarlane, M.D.), The Rutherford Foundation (PI: T. Cannon), and The Donaghue Foundation Early /Schizophrenia Initiative (PI: S.W. Woods).

The Authors

Tandy J. Miller, Ph.D., is Assistant Clinical Professor, Yale University, New Haven, CT. Thomas H. McGlashan, M.D., is Professor, Yale University, New Haven, CT. Joanna L. Rosen, PsyD, is at Yale University, New Haven, CT. Kristin Cadenhead, M.D., is at the University of California, San Diego, San Diego, CA. Tyrone Cannon, Ph.D., and Joseph Ventura, Ph.D., are at the University of California, Los Angeles, Los Angeles, CA. William McFarlane, M.D., is at the Maine Medical Center Research Institute, Portland, ME. Diana Perkins, M.D., is at the University of North Carolina, Chapel Hill, NC. Godfrey D. Pearlson is at Yale University, New Haven, CT. Scott W. Woods, M.D., is Associate Professor, Yale University, New Haven, CT.