Diagnosis and Screening for Psychotic Disorders in a Study of the Homeless

by Ezra S. Susser and Elmer L. Struening

Abstract

In a survey of homeless men, the authors found that screening scales for psychotic symptoms (Psychiatric Epidemiology Research Instrument) and signs (6-item scale of observational ratings) predicted a rating of psychosis (possible, probable, or definite) on a diagnostic interview (Structured Clinical Interview for DSM-III-R: Psychotic Disorders) reasonably well, in a sample where psychosis was common. Although the two scales performed well when used in conjunction, neither scale showed adequate predictive power when used alone. The authors conclude that screening for psychotic disorders in community studies is feasible for some purposes. They suggest approaches to the use of diagnostic interviews and screening scales in future community studies that might enhance the interpretability of results as well as the efficacy of screening.

The question to which we attend in this article is the identification of psychoses in community studies using diagnostic assessments and screening instruments. There is a scarcity of published data pertaining to this issue, perhaps because psychoses are at low prevalence in community surveys. Since only a small number of psychoses are found in any given sample, it is difficult to assess the performance of diagnostic interviews and to measure sensitivity, specificity, and positive predictive value for screening scales. By comparison, diagnostic interviews and screening instruments used in community studies to detect depression and other common disorders have been much more widely tested and discussed (Weissman et al. 1977; Dohrenwend et al. 1980).

This article draws on findings from a study of homeless men in which the prevalence of psychosis was high (12 to 25 percent). We believe the study was unique in comparing the evaluation of psychotic disorders in a community sample by screening scales and by diagnosis in the same subjects, administered by the same interviewers and also by different interviewers. We discuss the reliability and validity of a diagnostic interview and of screening scales for psychotic ideation and behavior, the feasibility of screening for psychotic disorders, and the future use of screening instruments for psychotic disorders in community studies.

The homeless have unusual characteristics; some of the results may be generalizable to other nonpatient populations, but some may not. Nevertheless, the results should be relevant to research on schizophrenia and other psychoses among nonhomeless as well as homeless persons. Screening instruments that detect psychosis efficiently will give epidemiological studies of these conditions a new impetus.

Methods

In the spring and summer of 1985, we interviewed several samples of homeless men who were using the New York City public shelter system. A 52-page interview was administered to all, and additional questions were given to selected samples. This article is based on the "first-timer" sample of new arrivals.
at the men's shelter system, and on the "census" sample of men staying in the shelter system at the time. In the first-timer sample (n = 223), 53 percent of the men were black and 84 percent were under 40 years of age. Thirty-two percent were newly homeless, 42 percent had been homeless "now and then" before shelter entry, and the remainder had been long-term homeless. In the census sample (n = 695), 71 percent were black and 74 percent were under 40 years of age. Observations in the census sample were weighted because some shelters were oversampled. Further description of the samples can be found in other publications (Struening 1986; Susser et al. 1989).

The first-timers (n = 177) were given a lifetime diagnostic interview for psychiatric disorders involving psychotic symptomatology, the Structured Clinical Interview for DSM-III: Psychotic Disorders (SCID-PD; Spitzer and Williams 1985). The SCID-PD is derived from an early version of the Structured Clinical Interview for DSM-III-R (SCID; Spitzer et al. 1988), under development at the New York State Psychiatric Institute, and uses diagnostic criteria almost identical to DSM-III-R. The authors of the SCID assisted in adapting the diagnostic interview for this population. For instance, special attention was given to the detection of drug and alcohol intoxication or withdrawal as a cause of psychotic symptoms.

Seven interviewers (mainly graduate students in psychology and social work) were trained to use the SCID-PD. They initially received training from the authors of the SCID, and then attended several weeks of further training in diagnosis and interviewing with homeless persons, which included pilot work for the study. Throughout the study, the psychiatrist (E.S.S.) met regularly with the interviewers, individually and as a group; virtually every diagnostic decision was reviewed individually.

As a first step, the SCID-PD requires a decision on the presence of a lifetime history of psychosis, as defined in the DSM-III-R (American Psychiatric Association 1987, p. 404). When such a history is present, the interviewer proceeds to diagnosis. In our survey, any possibly relevant information was recorded, even when the interviewer did not think that there was a history of psychosis. In addition, interviewers were asked to document in narrative form any difficulties they had in reaching a diagnosis.

On review of the interviews by the psychiatrist, it became apparent that incomplete histories might have undermined the validity of the dichotomous assessment of psychosis (present vs. absent) and of specific diagnostic decisions. Therefore, all the SCID-PD interviews were later reanalyzed by the psychiatrist and rated "definite," "probable," "possible," or "negative" for psychosis. A "definite" rating required a fully elaborated symptom history or the direct observation of such florid symptoms as incoherence or delusional thinking. Symptoms must have been present for at least 2 weeks in the lifetime history and must have occurred independently of drug or alcohol intoxication or withdrawal. A "probable" rating was given when the respondent reported a history of psychotic symptoms or the interviewer observed a disturbance in reality testing, but there was still room for doubt. For example, the respondent answered "yes" to a question about auditory hallucinations but did not wish to elaborate, or had a thought disorder but was not markedly incoherent. A "possible" rating was given mainly when the respondent reported no psychotic symptoms, but history and observation left some evidence that there might be a history of psychosis. For example, there was a history of neuroleptic treatment and no alternative explanation for it; "negative" symptoms such as flat affect and poverty of speech were observed; or there was a report of abnormal perceptions that were not clearly psychotic (a possible rating was often based on more than one such finding in the same person). A "negative" rating meant there were no grounds to suspect a history of psychosis. A diagnosis for a psychotic disorder was retained only for those with a "definite" rating.

For the census sample, we lacked the resources to use a diagnostic interview. Instead, we used a 10-item scale adapted for use in the shelters from the Psychiatric Epidemiology Research Instrument (PERI; Dohrenwend et al. 1980; Shrout et al. 1988). Respondents were asked about selected psychotic symptoms over the past year. Unlike the usual procedure followed with the PERI in other studies, the men were asked not to report symptoms that occurred only while using drugs or alcohol. If the interviewer felt that a response reflected something other than a psychotic symptom, a qualification was noted: "cultural," "situational," "untruthful," or "other." For instance, if a respondent reported feeling that people wanted to harm him, but the interviewer judged that the response reflected vigilance in the unsafe social environment of the shelter rather than paranoid ideation, the qualification "situational" would be coded.

In both samples, we also used a 6-item scale of observational ratings
This small subset of behaviors associated with psychosis was designed to detect markedly deviant behavior that might reflect psychosis. For instance, respondents were rated for "unkempt or bizarre appearance" and for "extreme distrust or suspiciousness."

To permit comparisons of results from the diagnostic interviews and screening scales, both the SCID-PD and the two screening scales were used in selected subsamples. In a subsample of 31 men in the first-timer sample, the screening scales were completed by a shelter intake worker who had been trained by us in the use of the instruments; later on the same day, these same men were interviewed with the SCID-PD by a regular interviewer unaware of the screening scale results. To avoid biased selection of mentally ill first-timers into this subsample, the shelter intake worker was asked to add the psychotic screen to the routine interviews with new arrivals that were already part of his job. In a second subsample of 61 men in the first-timer sample, the interviewers first completed the screening scales as part of the standard survey interview and then themselves administered the SCID-PD; when the methods gave different results, interviewers commented on the reasons. These 61 men were not selected systematically from the first-timer sample; rather, in the second half of the study, interviewers were encouraged (but not required) to use both methods to facilitate comparisons. Perhaps as a result, men with psychotic disorders are under-represented among the 61 (see below). For the remainder of the 177 first-timers who received the SCID-PD, the OR but not the PERI was completed.

**Results**

We first present findings on the reliability and validity of the SCID-PD, the PERI, and the OR. A comparison between results on the SCID-PD and the screening scales follows.

**Reliability of the SCID-PD.** Interrater reliability was tested in 18 interviews; one interviewer interviewed, the other observed, and each recorded an independent assessment, without knowledge of the other's decision. Of the 18 respondents, 16 were part of the first-timer sample, and two were from a separate sample of men in the same shelter survey. The two interviewers recorded the same diagnostic decision for each of these 18 respondents, 4 of whom had a history of psychosis.

In a limited test-retest reliability study, 11 men staying in shelters were interviewed with the SCID-PD and then reinterviewed several weeks later by another interviewer who was unaware of the results of the first interview. Of the 11 reinterviews, 8 were by the psychiatrist (E.S.S.). Four men were given the same diagnosis in both interviews, two were given a diagnosis only in the first interview, and five were given a diagnosis in neither ($k = 0.65$).

**Validity of the SCID-PD.** The criterion for validity, as discussed here, is a diagnostic decision based on a thorough knowledge of symptoms and signs. With the use of longitudinal observation and of sources of information other than the psychiatric interview, that validity criterion can be measured, if only approximately. In the cross-sectional shelter survey, however, validity could not be measured directly. Instead, the relationship between the SCID-PD diagnosis and the validity criterion had to be inferred by the psychiatrist on the basis of discussions with interviewers and interviewers' narratives about incompleteness in the information elicited in the interviews.

Information was considered inadequate for a clear-cut decision as to psychosis, and therefore diagnosis, in 23 cases; 9 of these were rated as "probable" and 14 as "possible" for psychosis. Twenty-nine men were rated as having a "definite" psychotic history (14 had schizophrenia, 3 schizoaffective disorder, 3 depression with psychotic features, and 1 psychosis not otherwise specified). The interval between the frequency of "definites" (12 percent) and of "definites, probabilities, and possibles" (25 percent) estimates the uncertainty due to incompleteness in the histories.

The "definite" and "negative" ratings were judged to be based on sufficient information. The truth of that judgment could not be directly tested. The "definite" ratings, however, had face validity (see *Discussion*). To a lesser extent, the "negative" ratings also had face validity, since they were based on an extensive interview in which no evidence of a psychotic history emerged.

**Reliability of the PERI.** The internal consistency reliability of the 10 items in the PERI scale was high ($\rho = 0.87$). The "situational" factors often noted by interviewers for paranoia-like items (2 and 3 on Table 1) did not obscure the effect of psychotic ideation on scores for these two items. Corrected item-total correlations were lower on these items than for most other items but still above 0.5.
Table 1. Men in New York City shelters 1985: Qualifiers to items on the Psychiatric Epidemiology Research Instrument (PERI) scale for psychotic symptoms (n = 695)

<table>
<thead>
<tr>
<th>PERI item</th>
<th>n(%) with positive response (almost never, sometimes, fairly often, very often)</th>
<th>n(%) of positive responses Qualifiers coded by Interviewers</th>
<th>Not qualified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Heard noises/voices others can't hear</td>
<td>85(12)</td>
<td>6(7)</td>
<td>26(31)</td>
</tr>
<tr>
<td>2. Felt there were people who wanted to harm you</td>
<td>165(24)</td>
<td>3(2)</td>
<td>118(72)</td>
</tr>
<tr>
<td>3. Felt there was something odd/unusual going on around you</td>
<td>182(27)</td>
<td>6(3)</td>
<td>132(73)</td>
</tr>
<tr>
<td>4. Seen things others can't see</td>
<td>88(13)</td>
<td>15(17)</td>
<td>25(28)</td>
</tr>
<tr>
<td>5. Had special powers others don't have</td>
<td>84(13)</td>
<td>10(12)</td>
<td>20(24)</td>
</tr>
<tr>
<td>6. Thought you were possessed by spirit/devil</td>
<td>54(8)</td>
<td>12(22)</td>
<td>12(22)</td>
</tr>
<tr>
<td>7. Felt your thoughts were taken from you by some outside force</td>
<td>55(8)</td>
<td>6(11)</td>
<td>14(25)</td>
</tr>
<tr>
<td>8. Had ideas/thoughts nobody could understand</td>
<td>140(21)</td>
<td>6(4)</td>
<td>45(32)</td>
</tr>
<tr>
<td>9. Felt thoughts were put into your head that were not your own</td>
<td>78(12)</td>
<td>7(9)</td>
<td>28(36)</td>
</tr>
<tr>
<td>10. Felt your mind was taken over by forces you couldn't control</td>
<td>57(8)</td>
<td>3(5)</td>
<td>13(23)</td>
</tr>
</tbody>
</table>

1 Possible responses were: never = 0; almost never = 1; sometimes = 2; fairly often = 3; very often = 4. Items are abbreviated for table.

Validity of the PERI. The qualifications coded by interviewers quantified their impressions as to whether positive responses ("almost never," "sometimes," "fairly often," "very often") to PERI items represented psychotic ideation or other factors. Table 1 shows the number of positive responses for each item and the proportion of the positive responses that were qualified. For several items, the majority of positive responses were qualified. For instance, for item 2 on paranoid thoughts, 72 percent of positive responses were qualified as "situationals," indicating that the interviewers usually perceived the respondents to be reporting the real dangers of life in the shelters rather than paranoia. Similarly, for item 3 on odd goings-on, 73 percent of positive responses were qualified as "situationals." For some items, the milder positive responses "almost never" and "sometimes" were qualified frequently, while the stronger response "very often" was not.

For several PERI items, mean scores in our sample were lower than mean scores of men in a community study in north Manhattan (Shrout et al. 1988) (table 2). The overall mean for the sum of scores on the 10 items was actually higher in the community sample (4.0) than in the shelter sample (3.3). Yet there is much evidence that psychosis should be more prevalent in the shelter sample—for instance, the high frequency of prior psychiatric hospitalization. The scale seems to have behaved differently in the study of the homeless compared with the study of north Manhattan. However, as noted above and discussed later, it was not used identically in the two studies.
Table 2. Comparison of findings in men in New York City shelters 1985 (n = 695) and north Manhattan community sample (n = 111)\(^1\) on PERI scale for psychotic symptoms

<table>
<thead>
<tr>
<th>PERI item(^2)</th>
<th>Never (0)</th>
<th>Almost never (1)</th>
<th>Sometimes (2)</th>
<th>Fairly often (3)</th>
<th>Very often (4)</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Heard noises/voices others can't hear</td>
<td>Community 83(75)</td>
<td>18(16)</td>
<td>8(7)</td>
<td>0(0)</td>
<td>2(2)</td>
<td>0.378</td>
<td>0.775</td>
</tr>
<tr>
<td></td>
<td>Shelter 597(88)</td>
<td>14(2)</td>
<td>44(6)</td>
<td>10(1)</td>
<td>17(2)</td>
<td>0.287</td>
<td>0.841</td>
</tr>
<tr>
<td>2. Felt there were people who wanted to harm you</td>
<td>Community 70(63)</td>
<td>22(20)</td>
<td>16(14)</td>
<td>1(0.9)</td>
<td>2(2)</td>
<td>0.585</td>
<td>0.893</td>
</tr>
<tr>
<td></td>
<td>Shelter 516(76)</td>
<td>19(3)</td>
<td>110(16)</td>
<td>17(2)</td>
<td>19(3)</td>
<td>0.523</td>
<td>1.016</td>
</tr>
<tr>
<td>3. Felt there was something odd/unusual going on around you</td>
<td>Community 56(50)</td>
<td>24(22)</td>
<td>24(22)</td>
<td>2(2)</td>
<td>5(4)</td>
<td>0.883</td>
<td>1.093</td>
</tr>
<tr>
<td></td>
<td>Shelter 499(73)</td>
<td>12(2)</td>
<td>117(17)</td>
<td>26(4)</td>
<td>27(4)</td>
<td>0.622</td>
<td>1.124</td>
</tr>
<tr>
<td>4. Seen things others can't see</td>
<td>Community 98(88)</td>
<td>7(6)</td>
<td>6(5)</td>
<td>0(0)</td>
<td>0(0)</td>
<td>0.171</td>
<td>0.502</td>
</tr>
<tr>
<td></td>
<td>Shelter 592(87)</td>
<td>16(2)</td>
<td>48(7)</td>
<td>13(2)</td>
<td>11(1)</td>
<td>0.279</td>
<td>0.798</td>
</tr>
<tr>
<td>5. Had special powers others don't have</td>
<td>Community 88(79)</td>
<td>11(10)</td>
<td>6(5)</td>
<td>3(3)</td>
<td>3(3)</td>
<td>0.396</td>
<td>0.917</td>
</tr>
<tr>
<td></td>
<td>Shelter 597(87)</td>
<td>9(1)</td>
<td>39(6)</td>
<td>11(2)</td>
<td>25(4)</td>
<td>0.321</td>
<td>0.932</td>
</tr>
<tr>
<td>6. Thought you were possessed by spirit/devil</td>
<td>Community 108(97)</td>
<td>1(0.9)</td>
<td>2(2)</td>
<td>0(0)</td>
<td>0(0)</td>
<td>0.045</td>
<td>0.282</td>
</tr>
<tr>
<td></td>
<td>Shelter 628(92)</td>
<td>7(1)</td>
<td>31(5)</td>
<td>4(0.6)</td>
<td>12(2)</td>
<td>0.184</td>
<td>0.688</td>
</tr>
<tr>
<td>7. Felt your thoughts were taken from you by some outside force</td>
<td>Community 102(92)</td>
<td>4(4)</td>
<td>3(3)</td>
<td>1(0.9)</td>
<td>1(0.9)</td>
<td>0.153</td>
<td>0.591</td>
</tr>
<tr>
<td></td>
<td>Shelter 626(92)</td>
<td>7(1)</td>
<td>29(4)</td>
<td>5(0.7)</td>
<td>14(2)</td>
<td>0.197</td>
<td>0.730</td>
</tr>
<tr>
<td>8. Had ideas/thoughts nobody could understand</td>
<td>Community 59(62)</td>
<td>20(18)</td>
<td>15(13)</td>
<td>4(4)</td>
<td>3(3)</td>
<td>0.667</td>
<td>1.021</td>
</tr>
<tr>
<td></td>
<td>Shelter 541(79)</td>
<td>18(3)</td>
<td>85(12)</td>
<td>16(2)</td>
<td>21(3)</td>
<td>0.460</td>
<td>0.997</td>
</tr>
<tr>
<td>9. Felt thoughts were put into your head that were not your own</td>
<td>Community 81(73)</td>
<td>16(14)</td>
<td>9(8)</td>
<td>3(3)</td>
<td>2(2)</td>
<td>0.459</td>
<td>0.892</td>
</tr>
<tr>
<td></td>
<td>Shelter 602(88)</td>
<td>8(1)</td>
<td>45(7)</td>
<td>9(1)</td>
<td>16(2)</td>
<td>0.270</td>
<td>0.821</td>
</tr>
<tr>
<td>10. Felt your mind was taken over by forces you couldn't control</td>
<td>Community 91(82)</td>
<td>12(11)</td>
<td>6(5)</td>
<td>1(1)</td>
<td>1(1)</td>
<td>0.279</td>
<td>0.690</td>
</tr>
<tr>
<td></td>
<td>Shelter 624(92)</td>
<td>4(0.5)</td>
<td>33(5)</td>
<td>8(1)</td>
<td>12(2)</td>
<td>0.205</td>
<td>0.734</td>
</tr>
</tbody>
</table>

\(^1\) Data provided by Dr. B. Dohrenwend and Dr. P. Shrout, Social Psychiatry Research Unit, Columbia University.

\(^2\) Items from the Psychiatric Epidemiology Research Instrument (PERI) are abbreviated for table.
Reliability of the OR. The internal consistency reliability of the six items in the OR scale was adequate ($\alpha = 0.72$). Corrected item-total correlations were also adequate (table 3). The variance for most items was low, perhaps reflecting the fact that the items were designed to differentiate markedly deviant behavior from all other behavior, rather than to measure a continuum of behavior ranging from normal to abnormal.

Interrater reliability and test-retest reliability of the OR were not measured in the census sample. In a subsample of 31 first-timers, however, test-retest reliability between a shelter intake worker and SCID-PD interviewer could be measured. For the dichotomy < 12 versus $\geq 12$ (see below), $k$ was 0.52; of five disagreements, four were due to lower ratings by the shelter intake worker compared with the SCID-PD interviewer. The intraclass correlation type (2, 1) Shrout and Fleiss 1979 for the ordinal scale was 0.53.

Validity of the OR. The validity of the OR cannot be assessed by the criterion of completeness of information, since behavior observed in the interview is the only information relevant to the rating. Also, we did not require that interviewers systematically code their impressions as to whether observed deviant behaviors were a manifestation of psychosis. Therefore, we turn in this case to construct validity.

In the census sample, the OR showed a lower correlation with psychiatric hospitalization history (point-biserial correlation $\rho = 0.14$) than did the PERI (point-biserial correlation $\rho = 0.27$). Hospitalization history was uncommon in this sample (12 percent), so that the highest point-biserial correlation was much less than one. In the first-timer sample, however, where interviewers had clinical experience, the OR showed a high correlation with hospitalization history (point-biserial correlation $\rho = 0.56$). (The PERI was not used routinely in the first-timer sample.)

The OR also showed a lower correlation with a reported history of treatment with antipsychotic medication (point-biserial correlation $\rho = 0.14$) than did the PERI (point-biserial correlation $\rho = 0.27$). Again, the OR correlation with antipsychotic medication history was high for clinical interviewers in the first-timer sample (point-biserial correlation $\rho = 0.53$).

Observational ratings were not highly correlated with reported symptoms (OR-PERI product moment correlation $\rho = 0.23$).

Comparison Between Screening Scales and the SCID-PD. In the subsample of 31 men who received the SCID-PD (by interviewers unaware of the results from the OR and PERI), for a history of psychotic symptoms ratings were 4 "definite," 1 "probable," and 3 "possible." For statistical analyses, these 8 men were grouped together as SCID-PD "positive," and the remaining 23 men were counted as "negative." The OR and PERI, too, were dichotomized for the comparison between screening scales and the SCID-PD, since we were primarily interested in their potential to screen for psychosis.

On common sense grounds, we first tried a score of 12 or above for the OR dichotomy and 10 or above for the PERI dichotomy. Respondents who had every sign or symptom to a minor degree, half of them to a moderate degree, or a few to a great degree would score at or above these cutoff points. The cutoff points turned out to yield reasonable agreement and predictive power, and we shall use them here in comparing the screening scales with each other and with the diagnostic interview. Subsequent inspection of

Table 3. Men In New York City shelters (1985). Observational rating (OR) scale for psychotic behavior—Census sample ($n = 695$)

<table>
<thead>
<tr>
<th>OR Item</th>
<th>Mean</th>
<th>Corrected Item total correlation</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inappropriate affect</td>
<td>1.24</td>
<td>0.588</td>
<td>0.71</td>
</tr>
<tr>
<td>Unkempt or bizarre in appearance</td>
<td>1.24</td>
<td>0.510</td>
<td>0.66</td>
</tr>
<tr>
<td>So withdrawn he found it difficult to answer questions</td>
<td>1.35</td>
<td>0.469</td>
<td>0.80</td>
</tr>
<tr>
<td>Unusual ways of thinking about experiences</td>
<td>1.79</td>
<td>0.429</td>
<td>1.17</td>
</tr>
<tr>
<td>Apathetic or flat affect</td>
<td>1.26</td>
<td>0.518</td>
<td>0.72</td>
</tr>
<tr>
<td>Extreme attitudes of distrust/suspicion</td>
<td>1.29</td>
<td>0.347</td>
<td>0.74</td>
</tr>
</tbody>
</table>

Note.—$\rho = 0.72$.

1 Possible ratings were not at all (1), to a slight extent (2), to a modest extent (3), to a moderate extent (4), and to a large extent (5). Items are abbreviated for table.
the data revealed that alternative
cutoff points (> 10 or above on OR,
> 12 on PERI) would have led to
slightly better agreement within the
subsample, but we did not choose to
use them in the analysis. We were
especially concerned about the
specificity
of the OR and the sensi-
tivity
of the PERI, and the alternative
cutoff points would diminish both.
(In the subsample, the effect was
not evident because of small
numbers).

Again on common sense grounds,
as well as on the basis of prior epi-
demiological research, we expected
that many persons who had not
recovered from a disorder involving
psychosis would appear disturbed
(and be so rated on the OR) but
would not currently have symptoms
such as delusions or hallucinations
or would not acknowledge them (on
the PERI), while others would be
fairly normal in appearance and in
social interaction but would report
symptoms. In fact, the correlation
between the two scales was even
lower than anticipated, and neither
scale alone proved to be a good
predictor of the SCID-PD rating.
Therefore, we counted a respondent
to screen positive if he scored above
the cutoff point on
either the PERI or
the OR.

The agreement between the
screening scales and the diagnostic
interview, using this criterion, is
shown in table 4. As shown in
4b, there were only 9 SCID-PD
positive cases among these 61 men,
indicating a lower prevalence of
possible psychosis in this subsample
(15 percent), compared with both
the sample as a whole (25 percent)
and the subsample of 31 men
discussed above (26 percent). As
would be expected when prevalence
is lower, agreement was slightly less
(k = 0.57, positive predictive value
= 0.71), though still encouraging.

An alternative approach to the
comparison between methods is the
use of discriminant function
analysis. One such analysis yields a
linear combination of the two scale
scores as a means for classification as
psychotic or not; the linear function
selected is that which minimizes the
probability of misclassification. The
function derived from the subsam-
ple of 31 is shown in tables 5a and
5b; agreement (as measured by k)
between the screening scales and
the SCID-PD is somewhat less than
that achieved with the simpler
"either-or" method described
previously, although the positive
predictive value remains
high. However, in discriminant
analysis, the function that minimizes
misclassification will be linear only
under certain assumptions that do
not hold in our data—for instance,
that the variance in scale scores is
similar for those who are psychotic

Table 4. Men in New York City Shelters (1985): Comparison of
SCID-PD with the PERI and the OR screening scales

4a. Subsample of 31 men who were evaluated on SCID-PD by interviewers
unaware of results on screening scales

<table>
<thead>
<tr>
<th>SCID-PD</th>
<th>Screening scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Sensitivity = 0.75</td>
<td></td>
</tr>
<tr>
<td>Specificity = 0.91</td>
<td></td>
</tr>
<tr>
<td>Positive predictive value = 0.75</td>
<td></td>
</tr>
</tbody>
</table>

4b. Subsample of 61 men who received screening scales and SCID-PD
during the same interview

<table>
<thead>
<tr>
<th>SCID-PD</th>
<th>Screening scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Sensitivity = 0.56</td>
<td></td>
</tr>
<tr>
<td>Specificity = 0.96</td>
<td></td>
</tr>
<tr>
<td>Positive predictive value = 0.71</td>
<td></td>
</tr>
</tbody>
</table>

1 SCID-PD (Structured Clinical Interview for DSM-III; Psychotic Disorders) was counted "positive" for possible, probable, and definite ratings for psychosis; and "negative" for negative ratings.
2 Screening scales were counted "positive" for PERI (Psychiatric Epidemiology Research Instrument) scores > 10 or OR (observational rating) scores > 12; otherwise counted "negative."
Table 5. Men in New York City shelters (1985): Comparison of SCID-PD with the PERI and the OR screening scales. Discriminant function analysis

5a. Subsample of \( n = 31 \) men who were evaluated on SCID-PD by interviewers unaware of results on screening scales

Discriminant function derived:

\[
\text{SCID-PD} = \text{"positive" when } 0.16 \text{ PERI} + 0.62 \text{ OR} > 8.37 \\
\text{SCID-PD} = \text{"negative" when } 0.16 \text{ PERI} + 0.62 \text{ OR} < 8.37
\]

<table>
<thead>
<tr>
<th>SCID-PD</th>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening scales</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>23</td>
</tr>
</tbody>
</table>

5b. Subsample of 61 men who received screening scales and SCID-PD during the same interview

<table>
<thead>
<tr>
<th>SCID-PD</th>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening scales</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>52</td>
</tr>
</tbody>
</table>

1 The SCID-PD (Structured Clinical Interview for DSM-III: Psychotic Disorders) was counted "positive" for possible, probable, and definite ratings for psychosis; and "negative" for negative ratings.

2 Screening scales were counted "positive" or "negative" for the PERI (Psychiatric Epidemiology Research Instrument) and the OR (observational rating) according to the discriminant function.

Discussion

The SCID-PD. There was considerable uncertainty in the assessment of psychosis with the diagnostic interview. This was partly because respondents were frequently unable or unwilling to elaborate their history of treatment and of symptoms. A check on New York State records revealed that 25 percent of the men in the census sample who had a history of psychiatric hospitalization in a State facility did not acknowledge any psychiatric treatment history in the survey interview (Struening 1987). Others who reported concrete information, such as treatment history, were elusive in reporting symptom history. For instance, those who recalled that they had been treated with neuroleptic medications often could not recall the symptoms for which they were prescribed. As a result, we were sometimes uncertain whether an abnormal mental status was due to a nonpsychotic condition (e.g., schizotypal personality disorder) or reflected a residual phase of a psychotic disorder such as schizophrenia; similarly, in some cases, a history of treatment with antipsychotic medication might have been due either to nonpsychotic behavior (e.g., violent outbursts) or to psychosis.

Psychiatrists might have fared better than our interviewers in eliciting complete histories with the SCID-PD. Also, factors specific to a shelter survey, such as the lack of privacy and comfort in shelter interviews, may have inhibited respondents. Yet serious limitations of structured clinical diagnostic interviews in eliciting information on psychotic disorders have been documented even for psychiatrists interviewing samples of psychotic patients (Carpenter et al. 1976; Downing et al. 1980; Brockington and Meltzer 1982); the absence of longitudinal observation and of other sources of information (for instance, interviews with relatives and staff, and hospital records) constrains validity. These limitations are especially apparent in the nonacute stage of a psychotic disorder—for instance, in a followup study—because symptoms tend to be less florid, and patients may have learned to deny them (World Health Organization 1979, p. 390). Further uncertainty in the history should be expected in community...
studies, where one encounters relatively less severe disorder so that cases tend to be less clear-cut (Dohrenwend et al. 1978); where psychosis, when present in the history, is less likely to be current; where respondents have had less recent practice answering questions about psychotic symptoms; and where the distinction between a clinical situation and a research interview may be sharply perceived by the respondent, inhibiting disclosure of symptoms to the research interviewer (Pulver and Carpenter 1983). Thus, even with more highly trained interviewers, a better controlled interview situation, and a less elusive study population, information might often be inadequate for a clear-cut decision on history of psychosis and diagnosis in a community study.

Our strategy for analyzing and reporting results from the diagnostic interview may be useful in other community studies when information is incomplete for many respondents. A rating for psychosis (definite, probable, possible, or negative) can be made before diagnosis. The "definite" rating, given only when the history of psychosis is well elaborated or the signs are florid, can then be considered a valid indicator of psychosis.1 A person with no psychotic history will not be, except in cases of the most precise and unlikely mimicry, rated "definite." It can be inferred that among the homeless men we surveyed, the true lifetime prevalence of psychosis lies in between the frequency of "definites" (12 percent) and that of "probables," "possibles" (25 percent).

While simple to adopt, this approach could add substantially to the interpretability of reported results for psychotic disorders in community studies, including some family studies. It maintains a needed focus on validity as well as reliability (Carey and Gottesman 1978). For current psychosis, the widely used Present State Examination (PSE; Wing et al. 1974) already allows for systematic recording of uncertainty in rating signs and symptoms and in judging psychosis; for lifetime psychosis, a systematic assessment of uncertainty may be even more necessary.

The PERI. Although the scale as a whole proved reliable, interviewers often doubted the validity of positive responses to individual items. While these subjective judgments by lay interviewers cannot be considered a validity criterion, they do indicate that responses to PERI items may have commonly been affected by extraneous factors. Interestingly, investigators screening for psychosis with similar questions in African countries have reported similar results. In Guinea-Bissau and in Ethiopia, respondents who reported that someone might want to harm them were often not psychotic; their positive response was thought to be due to cultural or situational factors (de Jong et al. 1986; Kortmann 1987). Possibly the distinct life circumstances and subculture of homeless persons lead to difficulties that resemble those encountered in using such questions in African countries. Alternatively, the problems may apply to other samples in the United States but have not yet been identified and reported. A recent report based on Epidemiologic Catchment Area data suggests that in an immigrant group in Los Angeles, in which many persons are not legal residents of the United States, some persons reported having believed that "someone was watching or spying on you" (a Diagnostic Interview Schedule [DIS; Robins et al. 1981] item) for legitimate reasons (Rubio-Stipec et al. 1989).

It is somewhat perplexing that mean scores for several PERI items were higher in a north Manhattan community study than in our sample. In fact, the distribution of responses for the two samples differed. On the PERI scale, a greater proportion of subjects in the shelter sample did in fact register very high scores (> 12) compared with the community sample. At the same time, a higher proportion registered very low scores. The different distribution is also observed in responses to individual PERI items (table 2); for several items, shelter respondents were notably less likely than community respondents to use the "almost never" response. Among the possible causes for the different distribution of responses in the homeless sample are: the deliberate exclusion of psychotic symptoms attributed by the respondent to drug or alcohol use in the

1"Definite" ratings may be valid even if they often do not replicate on a single reinterview; since psychotic symptoms are often forgotten or denied, persons with a psychotic history may not reveal it with consistency. A valid "definite" rating should still prove reliable, however, when one repeats a sequence of interviews rather than a single interview.
Feasibility of Screening for Psychosis. The degree of agreement found between these two simple screening scales and the diagnostic interview (κ = 0.66 and 0.57, respectively, in the two subsamples) is encouraging. While the results require replication in larger samples, they suggest that despite the daunting problems of measurement discussed above, adequate agreement between these methods is an achievable goal. No doubt the use of a general rating for psychosis, instead of a specific diagnosis, enhances agreement, and the rather high prevalence of psychosis in the sample also provides an advantage over other community studies (Dohrenwend et al. 1978). On the other hand, the methods differ more than is usual in such comparisons; for instance, the point-biserial correlations between most behavioral rating scales and psychiatric hospitalization history were extremely low (0.03 to 0.06).

The most challenging problems with the use of the OR arose as a result of the differences between shelter and nonshelter homeless persons. Interviewers familiar with both settings quickly recognized that different norms applied to these two groups. Thus, a rather disheveled man wearing a coat in mid-summer would be considered “unkempt or bizarre in appearance” if he had been staying for some time in the shelter; however, if he had been staying in the street recently, the coat could be attributed to his need for a blanket at night, and his disheveled appearance to lack of access to showers. Furthermore, since many homeless men dwell episodically in the shelters and at other times “sleep rough,” the appropriate reference group was not always clear.

Thus, agreement would have been considerably better if the diagnostic assessment had been restricted to current psychosis, as is the practice for the PSE. Also, improvements in the wording of some PERI items might have reduced false positives due to high PERI scores. We infer that screening for current psychotic disorders is indeed feasible, and that the predictive power of screening results reported here could be improved upon in future studies.

We now consider the circumstances in which screening for psychosis might be useful. The usefulness of the PERI and OR screening scales in identifying cases of psychosis must be judged with reference to the purpose of the screening. In the shelter survey, there were three purposes in mind: (1) The screening scales were used in the survey to yield a crude estimate of the prevalence of psychosis (Susser et al. 1989). (2) We
sought to develop a means for shelter workers to identify persons with psychotic disorders who might be in need of referral to treatment. (3) We wanted to examine the feasibility of using screening scales for psychosis in other community studies. We restrict comment here to the third issue.

Screening questions for psychosis have been used in some community studies, both in developing and in developed countries. The findings presented here suggest that further exploration of their use in community prevalence studies would be worthwhile. Shrout et al. (1986) have discussed the use of screening scales in a two-stage study to assess the prevalence of a psychiatric disorder in the community as an alternative to the usual one-stage community survey. In the first stage, a screening interview is administered to a representative sample; in the second stage, a fraction \( f_2 \) of those who screen positive and a fraction \( f_1 \) of those who screen negative are reinterviewed with a diagnostic interview. The information from the two stages can be combined to yield an estimate of prevalence.

The cost-effectiveness of the two-stage approach is greatest when the prevalence of the disorder is low, the cost of the screen is minimal compared with the diagnostic interview, and the sensitivity and specificity are high (Shrout, in press). For the assessment of psychosis in community surveys, the disorder is rare and the cost of administering these or similar screening scales is minimal compared with a diagnostic interview. Our findings suggest, in addition, that the sensitivity and specificity of screening scales may be acceptable, especially for current as opposed to lifetime psychosis. Therefore, the use of screening scales may turn out to be highly cost-effective.

The use of screening scales might also be considered in community studies that aim to identify persons in need of further evaluation or treatment for psychosis. In some studies, one may be concerned to identify almost all cases (high sensitivity), while in others one may aim to identify only the majority of cases while minimizing false positives (high positive predictive value). We suggest that screening scales for psychosis are best suited to the latter purpose. For psychosis, high sensitivity is difficult to achieve without lowering specificity to unacceptable levels. Yet a good positive predictive value may be achieved in settings where the prevalence of psychosis is relatively high—for instance, in some general practice and medical emergency room settings, in jails, and among the homeless. In the shelter survey, the results from the comparison of the screening scales with the SCID-PD were consistent with this view; however, there was no final criterion for the presence of psychosis, so the point is still not proven.

Whatever the purpose of the screening, the sensitivity and specificity of these or similar screening scales evidently need to be retested in a subsample within each study. The pattern of responses to PERI items varies across populations, and OR ratings vary by experience and training of the interviewers. Populations differ, too, in the type and severity of psychotic disorders. These and other factors will cause variation in sensitivity and specificity.

Our findings also bear on the choice and construction of screening scales for psychosis. When screening scales are used, it seems advisable to measure both reported symptoms and observed behavior, because one cannot assume a high correlation between the two. Behavioral ratings are probably less reliable and less valid than reported symptoms. Nevertheless, in a prevalence study, they may be needed to raise sensitivity, since they identify a significant number of cases that would otherwise be missed; in our sample the use of a dichotomized OR scale did not undermine specificity. In attempts to screen for evaluation and treatment, behavioral ratings may be needed to identify an important subgroup of psychotic persons who are severely disturbed but who cannot acknowledge their symptoms in an interview. Unfortunately, most commonly used screening scales that ask about psychotic ideation do not supplement self-report with behavioral ratings.

In a discussion of screening for psychiatric disorders, Katschnig (1987, p. 19) stated:

Possible [psychotic] patients may be categorized into three groups: those presenting florid and overt psychotic behavior which can be readily recognized also by non-psychiatrists; those not exhibiting overt psychotic behavior, but after

---

Anthony et al. (1985) have reported that among 11 community respondents with schizophrenia diagnosed by clinical interview but "missed" by the DIS, only 2 reported no psychotic symptoms at all in the DIS interview; DIS interviewers' marginal notes recorded bizarre or threatening behavior for both. Wittchen et al. (1985), similarly, reported that clinicians using the DIS made observations of psychotic symptoms in several former psychiatric inpatients who had a history of florid psychosis but nonetheless denied any history of psychotic symptoms in the DIS interview.
having been asked, admit psychotic experiences; and, finally, those not showing overt psychotic behavior but suffering from psychotic experiences, who do not admit the existence of the symptoms due to anxiety of stigma or of compulsory admission to a psychiatric hospital. The latter group constitutes a difficult problem not just in screening but also for experienced psychiatrists.

The usefulness of such a scheme, as a guide to the construction of screening scales, is borne out by our findings.

Conclusion

For psychotic disorders, opportunities to test and compare methods on large community samples with a high prevalence of psychosis are rare. Perhaps partly for this reason, methods for diagnosis and screening in community surveys have not yet been shown to be effective. For the DIS, for instance, studies conducted outside of treatment settings indicate that recall of lifetime symptoms of psychosis is limited (Pulver and Carpenter 1983; Wittchen et al. 1985) and that agreement with clinical interviews for the diagnosis of schizophrenia is poor (Anthony et al. 1985). In adding to the scant literature on methods of diagnosis and screening for schizophrenia and other psychoses in community surveys, we hope to contribute to the development of better epidemiological research on these disorders among nonhomeless as well as homeless persons.

References

American Psychiatric Association. 


Shrout, P.E., and Fleiss, J.L.
Intraclass correlations: Uses in assessing rater reliability. 


Struening, E.L. *A Study of Residents of the New York City Shelter System: Report to the New York City Department of Mental Health, Mental Retardation, and Alcoholism Services.* New York: New York State Psychiatric Institute, Epidemiology of Mental Disorders Research Department, 1986. (Copies available from author.)

Struening, E.L. *A Comparison of Evidence for a History of Disorder Derived From Survey and Record System Data.* New York: New York State Psychiatric Institute, Epidemiology of Mental Disorders Research Department, 1987. (Copies available from author.)


Acknowledgments

This work was supported in part by a Fellowship in Psychiatric Epidemiology USPHS grant MH-13043 from the National Institute of Mental Health, and by an award from the National Alliance for Research on Schizophrenia and Depression (NARSAD). We acknowledge the fiscal support of the following organizations and individuals: the New York City Department of Mental Health, Mental Retardation and Alcoholism Services, Dr. Sara Kellermann, Commissioner, grant 87206-88206; and the New York State Psychiatric Institute, Dr. Herbert Pardes, Director.

We thank Dr. Patrick Shrout for useful insights. We also thank Commissioner William Grinker and the staff of the New York City Human Resources Administration for their cooperation in the survey.

The Authors

Ezra S. Susser, M.D., M.P.H., is Research Assistant Professor, Statistical Sciences and Epidemiology Division, The Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, and the Department of Psychiatry, New York University School of Medicine, New York, NY. Elmer L. Struening, Ph.D., is Director, Epidemiology of Mental Disorders Research Department and Community Support Systems, New York State Psychiatric Institute, New York, NY, and Associate Professor, Columbia University School of Public Health, New York, NY.