Intervention Research in Psychosis: Issues Related to Clinical Assessment

by Delbert Robinson, Margaret Woerner, and Nina Schooler

Abstract

This article discusses efforts to optimize clinical assessment for intervention effectiveness trials. Generalizability is a crucial issue; investigators should choose their inclusion/exclusion criteria to increase subject inclusion and should collect data on the selection process to determine the extent of selection biases. Intervention research requires assessment instruments appropriate for a variety of treatment settings. We describe the Hillside Clinical Trials version of the Scale for the Assessment of Negative Symptoms, which has inpatient and outpatient versions to accommodate the different opportunities for social interaction in these settings. Lack of uniformity in assessment instruments complicates interpretation of results across studies and impedes communication of findings. We describe the 5-Dimensional Scale to Evaluate Psychopathology in Schizophrenia (5-STEPS), a collaborative effort to develop a standard change measure for schizophrenia treatment trials. We also discuss potential future strategies, including developing briefer yet reliable and valid diagnostic procedures, making trials more acceptable in a broad range of settings through the use of open-label treatment with blinded independent assessors, bridging efficacy and effectiveness designs by studying both a narrow efficacy and a broader effectiveness population simultaneously, and updating outcome domains to reflect current treatment strategies.

Keywords: Effectiveness research, assessment, diagnosis, generalizability, negative symptoms, 5-STEPS.


Assessment methodology will require modification as researchers move from relatively narrow-focus, short-term clinical trials studying treatment efficacy to studies of intervention effectiveness using broad-based sampling and long-term designs. A further impetus to change is the requirement by the National Institute of Mental Health (Department of Health and Human Services [DHHS] 1998) that specialized mental health intervention research centers "increase awareness of the importance of research for addressing important public health issues of patient care and clinical treatment" through "a commitment to outreach, patient education, public information and professional education."

In this article we will focus on completed or ongoing work in areas that are relevant to assessment for intervention studies: generalizability, modification of instruments to assess patients during long-term trials, and development of standard instruments to facilitate research and communication of research findings to the nonresearch community. We will also discuss areas that may be developed in the future.

Generalizability of Study Findings

A fundamental question in evaluating any research is, How generalizable are the findings to the overall population of patients? Systematic biases in research sample selection may affect overall study outcomes, limit the ability to extrapolate research findings to clinical settings, and ultimately limit the knowledge about effective treatment of groups who are underrepresented in research samples. The recognition that women and minority groups have been underrepresented in much medical research led the National Institutes of Health to make special efforts to correct the problem (DHHS 1994). Generalizability is a central question for any treatment study but is especially important in an intervention context.

To what degree have selection biases been present in mental illness treatment research? There are few data on
selection biases in psychosis treatment trials, and very few studies have compared the characteristics of patients entering a study with the characteristics of subjects in the overall population sampled. Two studies (Carr and Whittenbaum 1968; Schubert et al. 1984) of subject selection factors found that patients with schizophrenia are more likely than patients with other psychiatric diagnoses to refuse to participate in research. Thus, selection biases may be of particular concern for studies of patients with psychotic disorders.

Our investigation of this issue employed data from the Treatment Strategies in Schizophrenia (TSS) study (Schooler et al. 1997), in which all inpatients with clinical diagnoses of schizophrenia or schizophreniform or schizoaffective disorder at five sites varying in clinical organization and patient population were screened. There were 6,012 patients screened for the study, and only 528 (9%) entered (Robinson et al. 1996). This degree of selectivity certainly suggests that selection biases need to be investigated.

Systematic selection biases may result from specific study inclusion and exclusion criteria as well as refusal on the part of eligible patients to participate in the study. Investigators have the ability to decrease selectivity by broadening their inclusion and exclusion criteria. When designing studies, investigators should carefully assess whether a criterion is needed for scientific or human subjects reasons; often, criteria are copied from one protocol to another without reevaluation. Even with careful design, investigators may inadvertently introduce selection biases by their choice of inclusion and exclusion criteria. For example, among the TSS criteria (Robinson et al. 1996), only one, current pregnancy, was directly related to gender. This criterion excluded only 59 screened women. Other criteria excluded more women; the requirement that patients have weekly contact with their family of origin (necessary for the family therapy intervention in the study), for instance, excluded 1,103 women, 45 percent of screened women. Thus, investigators should monitor screening to detect unintended biases introduced by their criteria.

The consent process may introduce biases that are not under the control of investigators. Both the TSS study (Robinson et al. 1996) and an earlier study by Spohn and Fitzpatrick (1980) (n = 195) found that consenting subjects were younger, better educated, more likely to be African-American, and less likely to be women than were eligible subjects who refused study participation. The extent of agreement between the studies, despite considerable differences in the requirements of the research projects whose recruitment was investigated, suggests that there are some influential factors underlying the willingness to participate in research.

What can investigators do with information they collect on selection biases in their studies? By knowing the extent of sampling bias and making estimates of the response rates of the populations not entered, one can model the results of a study performed with unbiased samples. For many studies, modeled study results will be substantially different from actual results only if selection biases are very large (Robinson et al. 1996).

Until recently, researchers’ ability to model the effects of selection biases on study results has been limited by a lack of knowledge about the outcomes of patients who do not enter studies. Investigators were forced to use assumptions of response rates from other studies or “worst case scenarios.” To comply with requests from managed care organizations, many hospitals are now employing systematic clinical outcome assessments for all patients as part of their quality improvement efforts. This trend will provide better estimates of the response rates for subjects who do not enter trials, and these estimates may be used in models of the effects of selection biases on study results.

Instruments

Instrument development for intervention trials can draw upon past extensive efforts; Thornley and Adams (1998) counted 640 instruments in their review of 2,000 controlled treatment trials in schizophrenia. This section focuses on two areas of work for instrument modification and development for intervention studies: instrument modifications for long-term studies and standard change instruments for use by the field.

Instrument Modifications for Long-Term Studies. It is common for patients to change treatment settings during the course of long-term studies; a patient who does well may progress from inpatient treatment to partial hospitalization to community-based treatment. While the assessment of many symptoms is not affected by the treatment settings (e.g., positive symptoms may be assessed the same way for inpatients and outpatients), assessment of some other important symptoms and other outcome variables is influenced by the treatment setting. Thus, instruments may need modification for use in long-term trials. For example, we have developed a modification of the Scale for the Assessment of Negative Symptoms (SANS; Andreasen 1983) for use in clinical trials (for brevity, only some changes are discussed here; complete copies of the modified SANS are available upon request).

Our decision to modify the SANS rather than create a new instrument was based upon the value of the conceptual framework and content of the scale as well as its widespread use. The SANS was written with the expectation
Issues Related to Clinical Assessment

that investigators in general would use a 1-month time frame for rating, although other time frames could be used (Andreasen 1982). In most current clinical trials, a shorter interval change measure is desirable; our modification uses a 1-week rating time frame and revises anchors accordingly. In addition, to address the effects of changes in treatment setting, with concomitant differences in opportunities for social interaction, pursuit of leisure activities, and demonstration of “persistence” in role/occupational function, we developed parallel versions, one for use in inpatient, the other in outpatient settings.

An example of the issues involved in developing these parallel inpatient and outpatient versions is our adaptation, presented in the appendix, of the Impersistence at Work or School item of the original SANS. In our adaptation, inpatients have one expectation requirement for performance (i.e., the requirements of the inpatient unit), which they may fulfill to different degrees. Thus, in the inpatient version, subjects are rated for “participation in unit-appropriate activities.” In contrast, outpatients can attempt to function at many different levels, ranging from doing nothing to having full employment; at each level they may succeed to different degrees. Thus, we rate “role functioning” for outpatients first for the attempted role, then for how well the subject fulfills the attempted role. Without this inpatient/outpatient division, raters are confronted with making value judgments concerning the relative performance of someone who functions well at a low-expectation setting such as a day program versus someone who performs poorly in a high-expectation situation such as competitive employment. Rating both aspects of role functioning for outpatients enables one to consider each aspect separately when examining treatment effects.

Standard Change Instruments for Use by the Field.

Currently, three rating scales are widely used for the assessment of change in psychopathology in schizophrenia: the Brief Psychiatric Rating Scale (Overall and Gorham 1962; Woerner et al. 1988); the SANS and the Schedule for Assessment of Positive Symptoms (Andreasen 1983, 1984); and the Positive and Negative Syndrome Scale (Kay 1991). Each of these instruments has advantages and disadvantages and loyal users. But this use of different rating scales complicates the comparison of findings from study to study. Further, it compromises dissemination of findings to clinicians, policy makers, patients and their families, and other interested parties who have limited background, and interest, in scale differences.

Recognizing the need for a standard scale to measure change specifically geared to schizophrenia treatment trials, investigators from five research centers supported by the National Institute of Mental Health (University of Pittsburgh; Hillside Hospital; University of Iowa; University of Maryland; University of California, Los Angeles) and other institutions collaborated in developing such an instrument. The 5-Dimensional Scale to Evaluate Psychopathology in Schizophrenia (5-STEPs) is the result; it taps the following five dimensions of signs and symptoms in schizophrenia: delusions and hallucinations, disorganization, activation, depression-anxiety, and negative symptoms. The scale is currently being field tested.

Future Challenges in Intervention Research

Diagnostic Evaluation. Structured diagnostic instruments such as the Structured Clinical Interview for DSM-IV Disorders (SCID; First et al. 1994) are the norm for efficacy trials. However, they are time-consuming to administer, which limits their use in large sample studies and their relevance to “real-world” clinical care. Can researchers develop diagnostic procedures that are briefer yet sufficiently reliable and valid to meet scientific goals and good clinical practice? Doing so will be especially challenging since the populations in intervention trials will include patients with high rates of comorbid conditions. Accurate assessment of these conditions will be crucial for analyses examining differential treatment needs. A first approach to this problem may be to employ both clinician diagnoses and structured interview–based diagnoses in ongoing intervention trials and to develop data on their congruence (reliability) and comparative predictive value (validity).

Study Design. The “gold standard” for efficacy studies has been the randomized double-blind trial conducted at an academic center. What can be done to make intervention research feasible and acceptable in a broad range of settings? One approach is the use of randomized open-label treatment with blinded independent assessors as the raters. Although often employed in other contexts, independent assessors have not traditionally been used as the primary raters for pharmacological trials in psychotic disorders. One potential limitation of this design is that the independent assessor is cut off from contact with the clinical staff and thus deprived of important information about patient behavior. To compensate for this restricted information base, we have adopted the following procedures: (1) independent assessors complete baseline diagnostic assessments to increase their familiarity with patients and (2) weekly summaries of collateral information about patient behavior (e.g., attendance at activities) are prepared by clinicians and reviewed for objectivity, then given to the independent assessors to increase their access to relevant information.

The traditional model of the relationship between efficacy and effectiveness study designs is sequential—efficacy studies precede effectiveness studies. Once efficacy has been established, then it is time to eliminate...
restrictive exclusion criteria and design features and study treatments in real-world populations. Can researchers do effectiveness studies at earlier phases of testing of a treatment?

A possible strategy bridges efficacy and effectiveness designs by studying both the narrow efficacy population and the broader effectiveness population simultaneously. We have applied this model in a recently completed study that compared clozapine and risperidone (Schooler et al. 1999). The study had characteristics that are hallmarks of efficacy studies: subjects were randomly assigned to treatment, and treatment was administered under double-blind conditions. At the same time, the study had effectiveness study characteristics: a longer time frame (29 weeks) and broadened inclusion criteria that included subjects who were receiving other psychotropic medications, subjects who had comorbid substance abuse, and subjects who had prior exposure to the medications being investigated. Subjects were categorized as meeting “narrow” or “broad” inclusion criteria and were randomized to treatment within each group. Including both broad and narrow criteria in a study allows the explicit examination of whether outcome varies as a function of subject inclusion criteria.

**Updating Outcome Domains To Reflect Current Treatment Strategies.** The first round of studies with the new generation of antipsychotics compared them with standard antipsychotics; the next generation of studies will compare the new agents with each other. This will require modification of our assessment procedures, which were mostly developed for trials using conventional agents. New instruments will need to be developed; for example, side effect measures specific to the side effect profiles of the new agents must be constructed. Researchers may also need to change the focus of assessment batteries given the possibly different spectrum of action of the new generation of agents; more emphasis may need to be placed upon social adjustment, negative symptoms, and mood symptoms if these domains become more important targets for treatment.

**Conclusion**

The transition from treatment efficacy to intervention effectiveness trials requires researchers to reevaluate clinical assessment instruments and procedures. We have presented recent work on assessment for intervention trials as well as discussed areas for future development. New challenges will arise as intervention research in psychotic disorders matures. The recently introduced specialized mental health intervention research centers will offer additional help in meeting these challenges.

**References**


Andreasen, N.C. The Scale of the Assessment of Negative Symptoms (SANS). Iowa City, IA: University of Iowa, 1983.


Acknowledgments

The work presented in this article was supported in part by MH-41960 (Hillside Center for Intervention Research in Schizophrenia).

Investigators involved in developing 5-STEPS included Nina Schooler, John Kane, Margaret Woerner, and Delbert Robinson (Hillside Hospital); Nancy Andreasen and Michael Flaum (University of Iowa); William Carpenter and Robert Buchanan (University of Maryland); Stephen Marder and Joseph Ventura (University of California, Los Angeles); Jean Pierre Lindenmayer (Manhattan Psychiatric Center–Nathan Kline Institute for Psychiatric Research). Field testing of the scale is supported by a National Alliance for Research on Schizophrenia and Affective Disorders (NARSAD) grant to Nina Schooler.

The Authors

Delbert Robinson, M.D., is Assistant Professor of Psychiatry, Albert Einstein College of Medicine, Bronx, New York, and Director, Clinical Assessment and Training Unit, Hillside Center for Intervention Research in Schizophrenia, Hillside Hospital, North Shore–Long Island Jewish Health System, Glen Oaks, NY. Margaret Woerner, Ph.D., is Associate Professor, Department of Psychiatry, Albert Einstein College of Medicine, and Associate Director, Psychiatry Research, Hillside Hospital, North Shore–Long Island Jewish Health System. Nina Schooler, Ph.D., is Director, Psychiatry Research, Hillside Hospital, North Shore–Long Island Jewish Health System.

Appendix

Below are two items from the Hillside modification of the SANS. The first, Participation in Unit-Appropriate Activities, is from the inpatient version of the scale. Role Functioning, the corresponding item in the outpatient version of the scale, follows.

**Participation in Unit-Appropriate Activities**

Patients may have difficulty in attending and/or participating in assigned activities and general unit activities such as groups on the unit. Patients with mild impairment may attend activities but do not participate fully or do not complete assigned tasks. Patients with more severe impairment attend activities only with staff encouragement or not at all.

1. Participates appropriately in unit activities.
2. Mild. Patient requires some encouragement to attend or maintain participation in activities.
3. Moderate. Patient attends most activities but needs frequent prodding to attend or maintain participation.
4. Marked. Patient attends activities less than half the time and/or participates minimally.
5. Severe. Patient consistently fails to attend activities.

**Role Functioning**

The patient may have difficulty fulfilling social role expectations (e.g., employment, school, homemaking) as appropriate for his or her age and cultural background.

In rating role functioning, one must consider both (1) the difficulty of the role that the patient is attempting to fulfill and (2) how well the patient is functioning within that role. Therefore, this item is rated in two parts. First, the degree to which the patient's current role is appropriate to his/her age and social and cultural background is rated. Next, the degree to which the patient fulfills that role is rated separately.
Current Role Function—Level

Patient's current social/vocational level:

1. Age and socially appropriate role (full-time paid employment, matriculated in full-time school program NOT including psychiatric rehabilitation affiliated work or school programs, fulfills expectations of full-time homemaker, etc.).
2. As above not full-time (part-time student, part-time paid employment, etc.).
3. High-level psychiatric setting (high-level day program, vocational programs, etc.).
4. Low-expectation psychiatric setting (e.g., social/recreational programs or undemanding training programs).
5. Does not engage in any appropriate activities (no job, training program, or therapeutic program). Note: patients given this rating (a “five” on current Role Functioning—level) should in all cases be given a rating of “five” on the next item, current Role Functioning—Quality.

Current Role Function—Quality

Degree to which patient fulfills social/vocational role level recorded in the item above.

1. Fulfills expectations of current role (as rated in previous item).
2. Fulfills expectations of current role but with some difficulty (e.g., occasionally misses work, school or program without justifiable reason, occasionally fails to fulfill responsibilities).
3. Has definite difficulty fulfilling role responsibilities (e.g., consistently fails to attend and/or participate appropriately in current role).
4. Functioning at current role is seriously compromised and/or in danger of being dropped from current activity.
5. Not functioning in role.

Inpatients have one expectation requirement for performance (i.e., the requirement of the inpatient unit), which they may fulfill to different degrees. Thus, inpatients are rated for performance of role requirements based upon inpatient unit expectations, but not on role level.