Evaluating Clinical Trial Data: Outcome Measures

Robert R. Conley, M.D.

Schizophrenia is too often a severely disabling disease, and the discovery of interventions that can ease or eliminate symptoms without troubling side effects has long been the goal of schizophrenia research. In this endeavor, researchers, clinicians, and patients all desire an optimal outcome; outcome measures, which measure the relative success or failure of an intervention, are accordingly important. In addition, the costs of pharmacologic interventions—particularly of the atypical antipsychotics—in schizophrenia make the optimal measurement of treatment outcomes critical. Sometimes outcomes are focused on minimizing costly events, such as rehospitalization, rather than focusing on patient-oriented outcomes. This article discusses the outcome measures employed in 5 clinical trials comparing atypical antipsychotics, examining their usefulness and suggesting types of outcome measures that may be useful in the future. *(J Clin Psychiatry 2001;62[suppl 9]:23–26)*

Clinical outcomes measure the relative success or failure of an intervention. Schizophrenia is too often a severely disabling disease, and the discovery of interventions that can ease or eliminate symptoms without troubling side effects has long been the goal of schizophrenia research. In this endeavor, researchers, clinicians, and patients all desire an optimal outcome; outcome measures are accordingly important. In addition, the costs of pharmacologic interventions—in particular, the atypical antipsychotics—in schizophrenia make the optimal measurement of treatment outcomes critical. Sometimes outcomes are focused on minimizing costly events, such as rehospitalization, rather than focusing on patient-oriented outcomes. This article discusses the outcome measures employed in 5 clinical trials[1–5] comparing atypical antipsychotics, examining their usefulness and suggesting what outcome measures may be useful in the future.

One method of outcome measurement long in use is noting the change from baseline to endpoint in a traditional rating scale for schizophrenia, such as the Positive and Negative Syndrome Scale (PANSS) or the Brief Psychiatric Rating Scale (BPRS). More recently, outcome measures have identified the time to an event such as a relapse or rehospitalization, or the proportion of responders to an intervention in a trial. Time-to-event analysis estimates the probability that an event will occur at a given point in time. The most common type of time-to-event analysis used in the articles under discussion was the survival analysis, which estimates "the probability of survival as a function of time from a starting point, say, from the date of a diagnosis or of an intervention."[6(p137)] This is a welcome trend; when executed properly, such measures can be statistically powerful.

All the different outcome measures can be useful when used correctly. Changes on traditional rating scales have long been used in medication trials carried out to gain approval from the U.S. Food and Drug Administration. These traditional outcome measures should be combined with more recent innovations for complete analysis of the effect of a drug.

Outcome measures should aim to register optimal, "real world" changes in patients with schizophrenia, but the field of schizophrenia and psychosis research has until now done a poor job of describing good outcome in these patients. Conventional antipsychotics have certainly improved the lot of patients with schizophrenia in many ways, but the fact that these patients hardly lead normal lives remains an indictment of the field. Clinical trials evaluate the efficacy of an intervention, but clinical practice is often an effort to prevent negative outcomes or minimize certain costly events, such as an inpatient stay, rather than aiming at patient-oriented outcomes that emphasize the efficacy of the medication.

Part of the problem is the difficulty of defining and measuring outcomes. Schizophrenia is a heterogeneous illness with different domains of outcome. The question of effectiveness versus efficacy is an ongoing one. Another problem with studying outcome measures is that, when examining clinical trials, one frequently finds disparities between the study design, the stated hypothesis, and reported outcomes. This article deals with reported outcomes, with emphasis on their usefulness and how they differ from each other (Table 1).

From the Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, Baltimore.

Presented at the symposium "Evaluating Clinical Trial Data From Schizophrenia Research," which was held March 17, 2000, in Washington, D.C., and supported by an unrestricted educational grant from Janssen Pharmaceutica, L.P.

Reprint requests to: Robert R. Conley, M.D., Maryland Psychiatric Research Center, P.O. Box 21247, Baltimore, MD 21228.

Table 1. Outcome Measures in 5 Clinical Trials of Atypical Antipsychotics[a]

Outcome Measure	Tran et al[1]	Conley, Mahmoud, et al[2]	Ho et al[3]	QUEST[4]	Conley et al[5]
Abnormal Involuntary Movement Scale	✓				
Barnes Rating Scale for Drug-Induced Akathisia	✓		✓[b]		
Brief Psychiatric Rating Scale			✓		
Clinical Global Impressions scale				✓[c]	✓
Drug Attitude Inventory				✓	
EPS Checklist				✓	
Extrapyramidal Symptom Rating Scale		✓			
Global Assessment of Functioning					✓
Hamilton Rating Scale for Depression				✓	
Positive and Negative Syndrome Scale	✓	✓		✓	
Quality of Life Scale	✓[d]				
Scale for the Assessment of Negative Symptoms	✓		✓		
Scale for the Assessment of Positive Symptoms			✓		
Simpson-Angus Neurologic Rating Scale	✓		✓		
Proportion of patients with substantial EPS or medication change				✓	
Quality-of-life measures from PSYCH-BASE			✓		
Quality-of-life measures from CASH			✓		
Time to discharge					✓[b]
Time to rehospitalization					✓[b]

[a]Abbreviations: CASH = Comprehensive Assessment of Symptoms and History, EPS = extrapyramidal symptoms, PSYCHE-BASE = Psychiatric Status You Currently Have-Baseline Version.
[b]Brief Psychiatric Rating Scale derived from the Comprehensive Assessment of Symptoms and History.
[c]Proportion of patients improved "much" or "very much."
[d]By Kaplan-Meier method.

TRAN ET AL.

The data that researchers choose to report has a marked effect on outcome. The Tran et al.[1] study, for example, examined patients with schizophrenia, schizoaffective disorder, and schizophreniform disorder. A last-observation-carried-forward analysis was used for most change scores on the PANSS, the Scale for the Assessment of Negative Symptoms (SANS), and the Quality of Life Scale.[7] The study also reported the proportion of patients in whom more than 40% of symptoms abated. Any patient with schizophrenia whose symptoms are reduced by more than 40% on the PANSS has obviously done well, but one wonders, in this and other studies, how the figure chosen to report was arrived at and whether the results drove the reporting. One wonders, for example, if the largest proportion of patients had experienced a symptom abatement of 42% or 38%, whether that figure would have been reported.

Tran et al.[1] also looked at maintenance of response by the Kaplan-Meier method of survival analysis, but the results were affected by who was counted as a responder—only patients whose PANSS total scores improved 20% or more from baseline at week 8 were included in this analysis.

Tran et al.[1] defined side effects differently for the Simpson-Angus Neurologic Rating Scale, the Barnes Rating Scale for Drug-Induced Akathisia (BAS), and the Abnormal Involuntary Movement Scale (AIMS). New events were defined according to increasing scores on each scale; however, the argument can be made that these were not all new events but that some were incidents of worsening of continuing events, and continuing events will tend to regress toward the mean. The analysis was written in terms of the incidence of side effects, but the outcome measure recorded a score based on the number of categorical changes.

CONLEY, MAHMOUD, ET AL.

Conley, Mahmoud, et al.[2] compared risperidone and olanzapine in a double-blind, randomized trial in patients with schizophrenia and schizoaffective disorder. Outcome measures included the PANSS and the Extrapyramidal Symptom Rating Scale (ESRS). Currently available analyses are for patients completing the study—change in PANSS, change in ESRS, and proportion of responders by percentage improvement—but ultimately a last-observation-carried-forward analysis will be included.

Patients receiving either risperidone or olanzapine had improved PANSS total scores. In fact, there was little difference between the performance of each drug.

There are questions about the difference in positive symptom improvement—here expressed as proportions of responders by percentage improvement. The greatest difference between patients who responded while receiving risperidone and those who responded while receiving olanzapine is at the 40% responders level. If there is a normal distribution with this amount of mean modal response in a patient who was going to be a responder, it would be useful to know how to describe that distribution and that outcome.

HO ET AL.

Although Ho et al.[3] examined a convenience sample of patients in a clinic, the sample was well defined—only patients with schizophrenia were included, a fact that improved the generalizability of results. Ho et al. also examined change scores on the SANS, the Scale for the Assessment of Positive Symptoms (SAPS), the BPRS derived from the Comprehensive Assessment of Symptoms and History (CASH),[8] and quality-of-life measures from the CASH.

Outcome measures fluctuate. For example, the BPRS used in this study is derived from the CASH, yet that important distinction is often overlooked by readers who then use the derived BPRS to generalize from the results of such a study.

Ho et al.[3] also reported a response according to symptom dimensions. Symptoms were divided into 3 dimensions of psychopathology: The negative dimension was defined as the sum of alogia, anhedonia, avolition, and affective flattening global ratings in the SANS; the psychotic dimension was defined as the sum of the delusions and hallucinations global ratings in the SAPS; and the disorganized dimension was defined as the sum of the bizarre (disorganized) behavior, positive thought disorder, and inappropriate affect global ratings in the SAPS. These divisions were a positive step insofar as they recognized that there are different dimensions in schizophrenia. Nevertheless, the problem of defining these different dimensions and validating the definitions in order for outcome measures based on them to be useful is a continuing one.

QUEST

QUEST[4] also used an open-label, randomized design to study patients of all psychotic diagnoses; this fact is very important when considering the outcome measures. Outcome measures in this trial included the Extrapyramidal Symptoms Checklist,[9] the Hamilton Rating Scale for Depression (HAM-D), the Clinical Global Impressions scale (CGI), the PANSS, and the Drug Attitude Inventory (DAI-10). Using the HAM-D to assess a mixed population of patients with psychotic disorders is questionable on a number of grounds. In this case, it is hard to know whether the change registered in the HAM-D score occurred in patients with schizophrenia, schizoaffective disorder, or schizophreniform disorder. The researchers are to be commended, however, for their inclusion of the DAI-10, a patient-derived quality-of-treatment scale, among the outcome measures.

CONLEY ET AL.

Conley et al.[5] studied the outcome of patients in the Maryland State Mental Health database in an open-label, naturalistic study of rehospitalization rates of patients treated with atypical antipsychotics compared with patients treated with depot antipsychotics. Conley et al. looked at time to discharge and rehospitalization, time-to-event outcome measures, by the Kaplan-Meier method of survival analysis. It is currently difficult to assess how meaningful these events are. For example, time to discharge may be a meaningful outcome measure, or it may be driven mostly by reimbursement and rehospitalization factors.

Conley et al.[5] measured change in patients with the Global Assessment of Functioning (DSM-IV) and the CGI. They looked at outcomes by drug dosing and prescribing practices. Outcomes by drug dosing and prescribing practices are not rating scales, but if they can be used to produce changes, graphs, lines, and effect sizes, the results may be generalizable. Conley et al. plan, in future studies, to build change measures into their assessments, incorporating tests that are typically performed in the clinical environment.

In the time to rehospitalization analysis, patients receiving atypical antipsychotics were rehospitalized later than patients receiving depot formulations of typical antipsychotics. Since data for the patients receiving depot formulations were pooled, the reliability of this analysis might be questioned. This study[5] excluded patients receiving conventional antipsychotic medications orally to ensure that every patient received only 1 antipsychotic medication. The 1-year rate of recidivism for patients in the Maryland State Hospital release group is approximately 40%. All of the drugs tested registered rates of recidivism lower than 40%.

In another time-to-event measure, Conley et al.[5] considered time to discharge. They found that patients receiving risperidone were discharged faster than patients receiving olanzapine. One may ask whether these data are reliable or whether this difference relates to prescribing practices. Clinicians in Maryland have been systematically taught to administer risperidone in a tight dosage range, without titrating the medication very much. Olanzapine tends to be titrated considerably. This disparity in this measure could easily be a titration effect because clinicians have yet to discover an optimal dose of olanzapine. Time-to-event studies should be viewed with caution at this point.

DIRECTIONS FOR FUTURE STUDY

Many problems exist with current work in outcome measures. The studies under consideration here[1–5] had multiple types of assessments, and good corrections for the interrelation of the domains were rarely included. Under such circumstances, it is difficult to identify truly relevant outcomes. Answering different questions may require the use of different outcome measures. It is

especially important for us to differentiate hypothesis-generating experiments from clinical practice and enhancement experiments.

We fail to understand neurologic outcomes and how they affect clinical outcomes. We know, for example, how to prevent relapse in patients with schizophrenia early in the course of the disease with the use of benzodiazepines, a fact that has generated many clinically useful data, but this phenomenon has not been studied neurobiologically, and the question of how to combine the 2 aspects of study remains. Another problem in the field is that schizophrenia research is hampered by small sample sizes. Studies in cardiology or hematology, for example, routinely have sample sizes in the thousands, yielding great statistical power. As it is, researchers in schizophrenia must try to extrapolate results from hundreds of patients to thousands of patients, a problem that offers no easy solution but one we must solve as we identify outcome criteria, for the solution will ultimately drive our ability to generalize from our data.

The Human Genome Project, when complete, will offer opportunities in the field of outcome studies. The ability to identify genetic or surrogate markers for an illness such as schizophrenia has important implications for treatment, such as understanding the genetic factors that influence both therapeutic response and side effects and enabling clinicians to more accurately prescribe certain medications for specific patients.

Traditionally, efficacy has been measured in the clinical trial by physician-oriented outcome scales, such as the PANSS, measuring changes in patient psychopathology. Future studies should make greater use of alternative outcome measures such as time-to-event measures and, in particular, patient-oriented outcome scales. Patient satisfaction has a marked effect on both the ultimate efficacy of a drug and compliance rates. Indeed, Eisen et al.[10] studied whether psychiatric inpatients who completed the Behavior and Symptom Identification Scale, a self-report symptom and problem rating scale, felt more involved in the treatment process than patients who were treated according to standard protocols by either a psychiatric resident or an attending psychiatrist. Although treatment outcome did not differ among the groups, patients who completed the self-report form reported being handled with respect and dignity more often than those in the comparison groups. In addition, they assessed their involvement in treatment decision-making more highly than patients in either comparison group. These results suggest that the use of patient-oriented rating scales can have unexpected positive effects on the patient's sense of well-being.

CONCLUSION

In the future, we should examine how better to produce outcome measures such as time-to-event studies. When the goal is for psychotic patients to experience improvement in the form of a life event, a percentage change on a rating scale seems less and less like a meaningful measure of the improvement researchers, clinicians, and patients strive for. The meaningful change sought in schizophrenia research is an important event in a patient's life. We must ascertain how that event can best be defined, how it can best be measured, and how outcome measures can be made generalizable before they become truly useful.

Drug names: clozapine (Clozaril and others), olanzapine (Zyprexa), risperidone (Risperdal), quetiapine (Seroquel).

Disclosure of off-label usage: The author has determined that, to the best of his knowledge, no investigational information about pharmaceutical agents has been presented in this article that is outside U.S. Food and Drug Administration–approved labeling.

REFERENCES

1. Tran PV, Hamilton SH, Kuntz AJ, et al. Double-blind comparison of olanzapine versus risperidone in the treatment of schizophrenia and other psychotic disorders. J Clin Psychopharmacol 1997;17:407–418
2. Conley RR, Mahmoud R, and the Risperidone Study Group. Risperidone versus olanzapine in patients with schizophrenia and schizoaffective disorder. Presented at the 38th annual meeting of the American College of Neuropsychopharmacology; Dec 12–16, 1999; Acapulco, Mexico
3. Ho B-C, Miller D, Nopoulos P, et al. A comparative effectiveness study of risperidone and olanzapine in the treatment of schizophrenia. J Clin Psychiatry 1999;60:658–663
4. Mullen J, Reinstein M, Bari M, et al. Quetiapine and risperidone in outpatients with psychotic disorders: results of the QUEST Trial. Presented at the biennial meeting of the International Congress on Schizophrenia Research; April 17–21, 1999; Santa Fe, NM
5. Conley RR, Love RC, Kelly DL, et al. A comparison of rehospitalization rates between patients treated with atypical antipsychotics and those treated with depot antipsychotics. Presented at the 54th annual convention and scientific program of the Society of Biological Psychiatry; May 13–15, 1999; Washington, DC
6. Lang TA, Secic M. Assessing time-to-an-event as an endpoint: reporting survival analyses. In: How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers. Philadelphia, Pa: American College of Physicians; 1997:137
7. Heinrichs DW, Hanlon TE, Carpenter WTJ. The Quality of Life Scale: an instrument for rating the schizophrenic deficit syndrome. Schizophr Bull 1984;10:388–398
8. Andreasen N, Flaum M, Arndt S. The Comprehensive Assessment of Symptoms and History (CASH): an instrument for assessing psychopathology and diagnosis. Arch Gen Psychiatry 1993;49:615–623
9. Mintzer JE, Yeung PP, Mullen JA, et al. EPS variations in the elderly. In: New Research Abstracts of the 153rd Annual Meeting of the American Psychiatric Association; May 18, 2000; Chicago, Il. Abstract NR663:236
10. Eisen SV, Dickey B, Sederer LI. A self-report symptom and problem rating scale to increase inpatients' involvement in treatment. Psychiatr Serv 2000;51:349–353