Interpretations and Conclusions in the Clinical Trial

John P. Docherty, M.D.

Interpreting the results of a clinical trial and reaching conclusions are matters of assessing the integrity of the trial's design and methods and the validity of the findings and drawing inferences to make conclusions that will guide health care decisions; these decisions are the ultimate goal of the clinical trial. This article examines 5 clinical trials of atypical antipsychotic agents with particular attention to aspects of their validity and refers to other articles in this supplement when appropriate. Finally, the Behavioral Health Care Enterprise Model, designed to help systematize the process of drawing inferences and implications from these and other clinical trials and assessing their impact on various sectors of the health care system, is introduced. *(J Clin Psychiatry 2001;62[suppl 9]:40–43)*

Interpreting the results of a clinical trial can be thought of as a matter of assessing the validity of the findings and, from there, drawing logically sound and consistent inferences that will guide health care decisions, which is the ultimate goal of the clinical trial. This article examines 5 clinical trials of atypical antipsychotic agents with particular attention to aspects of their validity, reviews other articles in this supplement, and presents a model to help systematize the process of drawing inferences and implications from these and other clinical trials.

VALIDITY

The concept of validity, as it is applied to a clinical trial, is generally regarded as having 4 main aspects: internal validity, external validity, construct validity, and statistical validity (Table 1).

Internal Validity

Briefly, *internal validity* refers to the effectiveness of the design, the procedures, and the methods of a study to control potential confounds in drawing the conclusion that the experimental intervention is responsible for any differences observed between the control and experimental groups. Standard methods for "protecting" internal validity include the randomization of patients to treatment and "double-blinding" this standardization. Of the studies[1–5] under consideration here, 3—Tran et al.,[1] Conley, Mahmoud, et al.,[2] and QUEST[4]—were randomized, and 2—Tran et al.[1] and Conley, Mahmoud, et al.[2]—were blinded (Table 2). Randomization and blinding are discussed elsewhere in this supplement by Nina R. Schooler, Ph.D.[6]

External Validity

The safeguards of internal validity such as elimination of patients with comorbidity and restriction of the age range can work against external validity. The *external validity* or generalizability of a study is the extent to which "a study can produce unbiased inferences regarding the target population, beyond the subjects in the study."[7(p525)] To the extent possible, control of patient compliance, for example, is desirable from the standpoint of internal validity, that is, protecting the inference that the experimental drug is responsible for any observed differences between the treatment groups. However, control of patient compliance works against external validity because under real-world conditions, patient compliance tends to vary on the basis of various properties of a drug such as the number of doses per day and side effects. Therefore, a drug that is superior (more "effective") in the tightly controlled trial might be inferior (less "effective") due to compliance issues in a less controlled environment. A controlled psychosocial environment also supports internal validity by reducing variance in psychosocial variables that might affect outcome. Controlled psychosocial environments, however, are rare in the real world, and they limit the external validity of the clinical trial.

The use of concomitant medications can also have positive or negative effects on the validity of a trial. A protocol allowing the use of only a few adjunctive medications or none at all will enhance internal validity. Conversely, a

From Comprehensive Neuroscience, White Plains, N.Y.
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: John P. Docherty, M.D., Comprehensive Neuroscience, New York Hospital-Westchester Division, 21 Bloomingdale Road, White Plains, NY 10605.

Table 1. Validity

Type of Validity	Questions Addressed	Threats to Validity
Internal validity	To what extent can the intervention, rather than extraneous influences, be considered to account for the results, changes, or group differences?	Changes due to influences other than the experimental conditions, such as events (history) or processes (maturation) within the individual, repeated testing, statistical regression, and differential loss of subjects
External validity	To what extent do clinical trial conditions reflect the realities of clinical practice, ie, to what extent can the clinical trial results be generalized?	Clinical trial conditions that do not accurately reflect conditions of clinical practice, whether in population age or gender or in dosing and frequency, for example
Construct validity	Given that the intervention was responsible for change, what specific aspects of the intervention or arrangement were the causal agents, ie, what is the conceptual basis (construct) underlying the effect?	Alternative interpretations that could explain the effects of the intervention, ie, the conceptual basis of the findings, such as attention and contact with the subject, expectations of subjects or experimenters, cues of the environment
Statistical validity	To what extent can statistical tests be legitimately applied to the results?	Failure to record information and/or to interpret the results properly, ie, in ways that are consistent with the limitations of the data and sound statistical procedure

Table 2. Internal and External Validity in 5 Clinical Trials of Atypical Antipsychotics[a]

Characteristic	Tran et al[1]	Conley, Mahmoud, et al[2]	Ho et al[3]	QUEST[4]	Conley et al[5]
Randomization	Yes	Yes	No	Yes	No
Blinding	Yes	Yes	No	No	No
Adjunctive medication use	Reported	Not reported	Reported	Reported	Not reported
Medication compliance	Reported	Not reported	Not reported	Not reported	Not reported
Outpatient participation (no control of psychosocial environment)	Yes	Yes	No[a]	Yes	Yes

[a]Study included follow-up data obtained after discharge.

protocol designed to increase external validity might allow the unrestricted use of adjunctive medications. A range of patient-related variables may also have an impact on the generalizability of a study's findings. For example, the ratio of male to female patients affects external validity. As is pointed out by Samuel J. Keith, M.D., elsewhere in this supplement,[8] more men than women are generally recruited for schizophrenia clinical trials because men have an earlier onset of illness, poorer response to neuroleptics, and poorer outcomes.[9] This gender bias may limit the generalizability of study results from randomized patients to a wider population.[8] Three other key variables including the first episode (versus a chronic exacerbation), the relative severity of the acute episode, and the age of the patient can also affect external validity.[10]

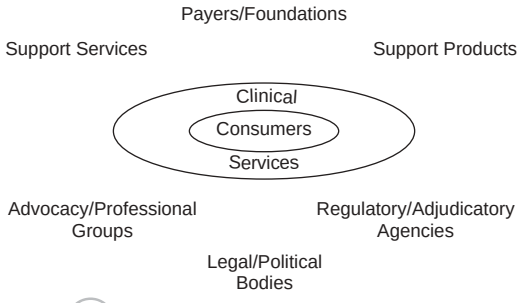
Construct Validity

Construct validity refers to the adequacy of the design, the methods, and the outcome measurements used to test the main hypotheses with conceptual integrity.

With regard to outcome measures, the concept of construct validity is pertinent to the selection of the primary outcome measurement and other assessment instruments employed in a particular clinical trial. Construct validity is a way of referring to the usefulness or adequacy of a particular test or measure as a reflection of a specific attribute or set of attributes. For example, is a clinician's global assessment of functioning an adequate measure of the degree of improvement a patient might experience following the treatment of a psychotic episode? The adequacy of a measure of change is typically assessed by a consideration of the convergent and discriminant validity of a scale, that is, the extent to which the scale agrees with other scales that have been developed as tests of the same attributes and differs from scales measuring a conceptually independent attribute or set of attributes. The essence of construct validity, however, refers to the adequacy of the measures to capture the key conceptually defined attribute. Thus, if "clinical outcome" is the concept and attribute to be measured, a variety of factors must be considered in assessing the adequacy of the measures of that attribute. Are the questions appropriate to the concept? Is the timing of the assessment appropriate to the concept? Does the measure assess the entire domain of the concept?

In terms of timing, an outcome measure applied at 1 week to an effect that conceptually is expected to take 3 to 6 weeks to occur would lack construct validity. Similarly, a particular assessment may measure only one aspect of a concept, so it should be determined whether the aspect being measured is conceptually appropriate to the treatment being administered. For example, a phobia can be conceptualized as a sense of subjective fear in the face of a particular object, a behavioral avoidance response in the face of a particular object, or physiologic arousal in the face of a particular object. Varying treatments may have selective

Figure 1. The Behavioral Health Care Enterprise Model



effects on each of these individual measures of phobia: biofeedback and relaxation therapy to reduce the physiologic response, cognitive therapy for subjective fear, and exposure therapy for behavioral avoidance. Correspondingly, in schizophrenia treatment a drug may differentially affect the relatively independent measures of outcome: symptoms and social and work function.

Statistical Validity

Statistical validity refers to the usefulness of an approach to data analysis to detect meaningful differences and not generate spurious differences. For example, the study by Tran et al.[1] lacks statistical validity because a 1-tailed test was used when assessing primary efficacy in the comparison of olanzapine with risperidone, lending in these circumstances to a classic error of rashness.[11] The necessary basis for considering this a valid test, such as a body of findings strongly suggesting that olanzapine would be found superior to risperidone in a clinical trial or vice versa, is lacking. The use of the 1-tailed test also entails an a priori determination that a finding that risperidone is superior to olanzapine would be of no interest to the investigator from a statistical standpoint.

THE BEHAVIORAL HEALTH CARE ENTERPRISE MODEL

The information derived from clinical trials finds use in a wide variety of settings. The Behavioral Health Care Enterprise Model allows for systematic and inclusive thinking about the usefulness of this information across the entire domain of behavioral health care (Figure 1). With regard to the payer sector, for example, if clear distinctions cannot be drawn among the medications, a payer can reasonably be expected to favor the least costly drug as the initial approach to care. At current prices, for effective levels of dosing, risperidone is less expensive than either olanzapine or quetiapine (Table 3). The data in Table 3 come from PriceProbe (First DataBank, San Bruno, Calif.) and National Disease and Therapeutic Index (IMS HEALTH, Plymouth Meeting, Pa.), March 2000.

Table 3. Comparative Daily Cost of Atypical Antipsychotics[a]

			Average Cost Per Day	
Drug	Dose, mg	AWP[b]	Any Indication	Schizophrenia
Olanzapine	2.5	$4.81	$9.81	$11.97
	5	$5.69		
	7.5	$5.69		
	10	$8.64		
	15	$12.93		
Quetiapine	25	$1.36	$6.88	$9.79
	100	$2.48		
	200	$4.68		
Risperidone	0.25	$2.53	$5.83	$8.26
	0.5	$2.53		
	1	$2.53		
	2	$4.22		
	3	$4.98		
	4	$6.56		

[a]Data from PriceProbe (First DataBank, San Bruno, Calif.) and National Disease and Therapeutic Index (IMS HEALTH, Plymouth Meeting, Pa.), March 2000; cost per day is calculated as (Sum[AWP price per tablet × number of tablets per day × number of mentions])/total number of mentions for product.
[b]Average wholesale price (AWP) is the price a retailer can expect to pay a wholesaler for any given product.

Support services include all of the activities that support clinical care—research, for example. Since strong differences among the atypical antipsychotics are unlikely to arise, research to establish any one atypical antipsychotic as the most effective has limited usefulness. Therefore, a more advantageous endeavor might be to establish if there are response differences to atypicals in patient subgroups. This use of research involves a process of logical inference. The quality and validity of this inference should be assessed by the integrity of its logical exegesis. Another use of research might be to attempt to determine genetic polymorphic predictors of individual response in both efficacy and toxicity to the different atypical antipsychotics.

In terms of *clinical services*, these data may help inform ultimate clinical decisions regarding which medication to select for an individual patient, how to use that medication, and what to tell the patient he or she might expect from treatment. Such information includes the comparative degree of clinical improvement achieved with available medications; the type, duration, and intensity of side effects; and the optimal range of medications. For example, from the group of studies discussed, there are 3 illustrative findings:

(1) Risperidone should be used for most patients in doses below 6 mg/day, so that efficacy is obtained but extrapyramidal side effects are minimized.[12]
(2) Olanzapine induces substantially greater weight gain than quetiapine or risperidone.[1,2,13–17]
(3) There is a suggestion, derived from 2 of the studies, that risperidone may have differential benefits in decreasing hallucinations and delusions.[2,3]

For *legal and political bodies*, the suggested evidence[5] of decreased relapse with the atypicals may give some

added support for increasing resources to support outpatient programs such as New York State's "Kendra's Law," legislation designed to ensure that potentially dangerous patients with mental illness are safely and effectively treated.[18] The atypical antipsychotics are both more benign and more effective than the conventional antipsychotics, which produced the noncompliance that led to this law. The main argument against outpatient programs has been the lack of compliance in seriously ill schizophrenia patients. But the enhancement of outpatient programs makes very good sense with the use of the atypical antipsychotics, given the improved side effect profiles, improved compliance, and decreased relapse rates demonstrated with these newer agents.

CONCLUSION

Examining issues of validity proves to be a helpful tool in interpreting the results of a clinical trial and drawing inferences that are useful for making conclusions. When assessing the data from 5 clinical trials of atypical antipsychotic agents with particular attention to aspects of their validity, the Behavioral Health Care Enterprise Model helps to systematize the process of developing conclusions from these clinical trials for each of the various sectors of the health care system.

Drug names: olanzapine (Zyprexa), quetiapine (Seroquel), risperidone (Risperdal).

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. Schooler NR. The statistical comparison of clinical trials. J Clin Psychiatry 2001;62(suppl 9):35–37
7. Last JM. A Dictionary of Epidemiology. 3rd ed. New York, NY: Oxford University Press; 1995. Cited by: Iverson C, Flanagin A, Fontanarosa PB, et al. American Medical Association Manual of Style: A Guide for Authors and Editors. 9th ed. Baltimore, Md: Williams & Wilkins; 1998
8. Keith SJ. Evaluating characteristics of patient selection and dropout rates. J Clin Psychiatry 2001;62(suppl 9):11–14
9. Lindamer LA, Lohr JB, Harris MJ, et al. Gender-related clinical differences in older patients with schizophrenia. J Clin Psychiatry 1999;60:61–67
10. Kane JM. Dose selection and comparator drugs in schizophrenia research. J Clin Psychiatry 2001;62(suppl 9):29–32
11. Greenhalgh T. How to read a paper: statistics for the non-statistician, I: different types of data need different statistical tests. BMJ 1997;315(7104): 364–366
12. Risperdal (risperidone). Physicians' Desk Reference. Montvale, NJ: Medical Economics; 2000:1453–1457
13. Beasley CM Jr. Safety of olanzapine. J Clin Psychiatry Monograph 1997; 15(2):19–21
14. Tran PV, Tollefson GD, Sanger TM, et al. Olanzapine versus haloperidol in the treatment of schizoaffective disorder: acute and long-term therapy. Br J Psychiatry 1999;174:15–22
15. Osser DN, Najarian DM, Dufresne RL. Olanzapine increases weight and serum triglyceride levels. J Clin Psychiatry 1999;60:767–770
16. Allison DB, Mentore JL, Heo M, et al. Antipsychotic-induced weight gain: a comprehensive research synthesis. Am J Psychiatry 1999;156: 1686–1696
17. Masand PS. Relative weight gain among antipsychotics [letter with reply]. J Clin Psychiatry 1999;60:706–708
18. Governor's "Kendra's Law" to protect public mentally ill. New York State Home Page. Available at: http://www.state.ny.us/governor/press/year99/May19_99.htm. Accessed September 12, 2000