

## Randomized Controlled Trials for Schizophrenia: Study Designs Targeted to Distinct Goals

T. Scott Stroup<sup>1,2</sup> and John R. Geddes<sup>3</sup>

<sup>2</sup>Department of Psychiatry, University of North Carolina, Chapel Hill, NC 27599-7160; <sup>3</sup>Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK

**Randomized controlled trials for antipsychotic drugs have a variety of design features suited to diverse purposes. Efficacy (or explanatory) trials seek to establish if a drug can reduce psychotic symptoms under ideal circumstances. To isolate drug effects, researchers enroll carefully selected patients. Specialized research personnel use rating scales of symptoms that are sensitive to drug effects as the study primary outcomes. Large simple trials (LSTs) are conducted at typical treatment settings with usual clinical personnel and enroll large numbers of participants so small but clinically important differences between treatment options can be detected. LSTs focus narrowly on clearly defined, patient-oriented outcomes. To some extent, practical trials can be conceptualized as hybrids of efficacy and large simple trials. Practical trials provide independent evidence to inform decision makers about the everyday effectiveness of clinically relevant alternative interventions. Practical trial researchers include a heterogeneous population of patients and collect data on a broad range of meaningful health outcomes at many types of practice settings intended to represent usual treatment. The designers of practical trials make trade-offs between internal validity, external validity, the breadth of issues addressed, and the ability to detect small differences. The different objectives of trials should be considered in the interpretation of the complete body of randomized evidence on antipsychotic drugs.**

*Key words:* randomized controlled trials/antipsychotic drugs/practical trials/large simple trials

### Introduction

Randomized controlled trials (RCTs) remain the gold standard for establishing evidence to support clinical and policy choices. RCTs comparing antipsychotic drugs,

however, have produced conflicting results and interpretations,<sup>1–3</sup> and many studies have been criticized for failing to address critical issues adequately.<sup>4–6</sup> This article aims to delineate the features of certain kinds of RCTs relevant to antipsychotic treatment of schizophrenia and to demonstrate that diverse trials may have complementary purposes that together can provide a broad understanding of antipsychotic drug effects.

Efficacy or explanatory trials are usually conducted by a drug's developer to meet requirements for marketing approval by drug regulatory agencies. These highly standardized, short-term studies compare a new drug with placebo or an active comparator or both and usually make no pretenses about a drug's long-term effectiveness. Safety, efficacy, and speed are critical priorities because companies seek rapid approval for a new drug.

Large simple trials (LSTs) are new for antipsychotic drugs and psychiatry but are widely used in other medical fields. These trials enroll great numbers of patients at typical practice settings and randomly assign treatments, but use few specialized research procedures and collect only minimal data focused on a critical outcome. Prominent LSTs in cardiology, eg, the Clopidogrel and Metoprolol in Myocardial Infarction Trial (COMMIT)<sup>7,8</sup> and Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT)<sup>9</sup> investigations, used mortality or stroke as primary outcomes. These influential studies have helped determine the relative benefits and risks of treatments for hypertension and acute myocardial infarction.

Practical or pragmatic trials are also relatively new to psychiatry, but several examples, including the recent Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE),<sup>10–14</sup> Rapid Tranquilization Clinical Trial (TREC),<sup>15–17</sup> and Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS)<sup>18,19</sup> investigations, have provided additional information about the relative risks and benefits of antipsychotic drugs. Practical trial designs can vary considerably in their complexity, but their overall goal is to provide comparative treatment information that can be applied broadly to relevant patients.

In this article, we describe the typical characteristics and goals of efficacy trials, LSTs, and practical trials and then use contemporary examples involving antipsychotic drugs

<sup>1</sup>To whom correspondence should be addressed; tel: 919-966-6846; fax: 919-966-7659, e-mail: sstroup@med.unc.edu.

**Table 1.** Typical Characteristics of Efficacy, Large Simple, and Practical Trials (Adapted from Stroup S. *Epidemiologia e Psichiatria Sociale*, vol 14, no 3, 2005)

	Efficacy Trials	Large Simple Trials	Practical Trials
Goal	Achieve regulatory approval to market drug	Compare treatment options to examine small but potentially important differences	To inform decision makers about clinical and policy choices
Specific Aims	To establish short-term efficacy and safety of a new drug	To determine comparative long-term safety or effectiveness	To examine relative benefits and risks of available treatments
Primary outcome	Improvement of target symptoms	A discrete and clinically meaningful outcome	A discrete and clinically meaningful outcome
Secondary outcomes	Safety measures; response rates	Few, if any	Many health-related outcomes
When conducted	Before a drug is marketed	Postmarket	Postmarket
Sponsor	Drug maker	Varies; drug maker, government agency, foundation	Varies; drug maker, government agency, foundation
Diagnostic criteria	Diagnosis by structured interview	Clinical diagnosis or structured interview	Clinical diagnosis or structured interview
Sample size	A few hundred	Thousands	Hundreds to thousands
Comparisons	Placebo and an active comparator	One or more active comparators; clinical equipoise	One or more active comparators; clinical equipoise
Dosing	Fixed dosing	Flexible dosing in clinically used range	Flexible dosing in clinically used range
Blinding	Double blinded	Open label, single, or double blinded	Open label, single, or double blinded
Duration	6 Weeks	1 Year or longer	1 Year or longer
Research sites	Experienced research sites capable of specialized procedures: number variable but usually less than 100	Sites are typical treatment settings: often hundreds of sites	Sites are typical treatment settings, may need capacity to measure many outcomes: often hundreds of sites
Research protocol	Strictly defined research protocol with frequent research assessments and procedures	Protocol very similar to usual practice; minimal additional assessments or procedures	Protocol designed to mimic usual decision making for antipsychotic drugs; many research procedures included
Comorbid medical and psychiatric conditions	Excluded to isolate the effects of the drug	Included except for specific contraindications to examine effects in typical patient populations	Included except for specific contraindications to examine effects in typical patient populations
Adjunctive and concomitant treatments	Excluded except specific drugs for limited time	Allowed as in usual practice	Allowed as in usual practice; some exclusions depending on study aims

to illustrate the merits, weaknesses, and what can be learned from each type of trial. Table 1 shows typical characteristics of these trials and Table 2 provides examples.

### Efficacy Trials: Can a Drug Work?

Efficacy (or explanatory) trials seek to establish if a drug *can* work under ideal circumstances.<sup>20</sup> Developers of new antipsychotic drugs use efficacy trials to establish a drug's capacity to reduce psychotic symptoms. The major goals in the design of efficacy trials is to make sure that any treatment effect is likely to be picked up and to reduce the background "noise" caused particularly by interrater variation in the assessment of outcomes and the presence of comorbidity, ie, to maximize the signal-to-noise ratio.

To achieve these aims, researchers conduct these studies in specialized settings using trained personnel and deploying rating scales of symptoms that are sensitive to drug effects as the study primary outcomes.<sup>20</sup> A purported antipsychotic is commonly compared both with placebo and with a known antipsychotic; this procedure helps to confirm "assay sensitivity", or that the study design and procedures are adequate to detect a difference between a drug known to be effective and placebo.<sup>21</sup>

Trained research personnel recruit patients experiencing an acute episode of schizophrenia and conduct specialized assessments to ensure that diagnostic and severity of illness criteria are met. Most individuals with medical illnesses, substance use disorders, or additional psychiatric diagnoses are excluded both to reduce the likelihood of

**Table 2.** Actual Characteristics of Representative Efficacy Large Simple and Practical Trials

	Efficacy Trials	Large Simple Trials	Practical Trials
Example Goal	Phase 3 US paliperidone trial <sup>22</sup> Achieve regulatory approval to market paliperidone	ZODIAC <sup>37</sup> Examine safety issue raised in premarketing studies	CATIE <sup>38</sup> To inform decision makers about relative effectiveness of antipsychotic drugs
Specific Aims	To determine short-term efficacy and safety of paliperidone	To determine long-term safety of ziprasidone	To examine relative benefits and risks of widely used antipsychotics
Primary outcome	Change in PANSS total score	1 Year nonsuicide mortality	Time to treatment discontinuation for any cause
Secondary outcomes	Many safety parameters; other clinical parameters	Causes of all deaths, hospitalization and reasons for hospitalizations, and treatment discontinuation	Reasons for treatment discontinuation; multiple other outcomes, including safety, costs, quality of life, and neurocognitive function
When conducted	Before marketing	Postmarket	Postmarket
Sponsor	Janssen	Pfizer	NIMH
Diagnostic criteria	Diagnosis by structured interview	Clinical diagnosis	Diagnosis by structured interview
Sample size	444	>18 000	1493
Comparisons	Paliperidone 6 or 12 mg daily vs olanzapine 10 mg daily vs placebo	Ziprasidone vs olanzapine	Olanzapine vs perphenazine vs quetiapine vs risperidone vs ziprasidone
Dosing	Fixed dosing	Flexible dosing in clinically used range	Flexible dosing in clinically used range
Blinding	Double blinded	Open label	Double blinded
Duration	6 Weeks	1 Year for primary outcome	18 Months
Research sites	74 Research sites with inpatient services in the United States	Almost 1000 sites in 18 countries: sites are typical treatment settings	57 sites in United States: Typical treatment settings capable of implementing detailed research protocol
Research protocol	Strictly defined research protocol with frequent research assessments and procedures	Protocol very similar to usual practice; minimal additional assessments or procedures	Protocol designed to mimic usual decision making for antipsychotic drugs; many research procedures included
Comorbid medical and psychiatric conditions	Substance dependence and medical conditions affecting drug metabolism excluded	Included except for specific contraindications	Included except for specific contraindications
Adjunctive and concomitant treatments	Antidepressants and mood stabilizers not allowed	Allowed as in usual practice	Allowed as in usual practice; other antipsychotics prohibited

serious adverse events and limit the variance in outcomes. To protect against ascertainment and performance biases, treatments are assigned randomly under double-blind conditions, so that neither the patient nor the research team providing clinical care and conducting outcome assessments know the treatment assignment. Typical efficacy trials are designed to follow the clinical course of participants for 6–8 weeks. Trained personnel conduct frequent assessments of symptoms and side effects using psychometrically validated instruments. The primary outcome is typically improvement on a symptom rating scale, such as the Positive and Negative Syndrome Scale (PANSS). Response, or improvement by some prespecified amount on a symptom rating scale, is a common secondary outcome. Known side effects and common laboratory parameters are measured systematically, while other adverse events are reported spontaneously if and when they occur.

Efficacy trials efficiently meet the requirements of drug-licensing agencies to show that a drug works and that it is safe to use. They work well for drug developers that want to complete trials rapidly, because the high costs of using many highly trained and geographically dispersed research centers can be offset against the reduced time to market and profitability. However, efficacy trials do not reveal much about how a new drug works in typical settings, with patients who may have medical and psychiatric comorbidities and who may take other medications. They do not tell us about long-term safety issues or about the effects of drug therapies on mortality, ability to work, or other issues that are important to patients.

#### *Example of an Efficacy (or Explanatory) Trial*

“Efficacy and safety of paliperidone extended-release tablets: results of a 6-week, randomized, placebo-controlled study” is a recent, typical example of an

efficacy trial.<sup>22</sup> Janssen (Titusville, NJ) conducted this phase 3 trial to meet regulatory requirements for approval of paliperidone as a treatment for acute schizophrenia. The trial compared 2 fixed dosages of paliperidone (6 and 12 mg daily) with placebo and olanzapine 10 mg daily for up to 6 weeks under double-blind conditions. Change in symptoms as measured by the PANSS total score was the primary outcome. Clinical response was defined as a 30% reduction in PANSS total score.

### *Eligibility*

Individuals aged 18 or older who were experiencing an acute episode of schizophrenia as indicated by a PANSS score of 70–120 (representing moderate to severe, but not mild or extremely severe, psychopathology) were eligible.<sup>22</sup> Patients met DSM-IV criteria for schizophrenia and agreed to at least 14 days of hospitalization. Extensive exclusion criteria included the following: substance dependence in the previous 6 months; medical conditions that might affect drug pharmacokinetics; history of tardive dyskinesia or neuroleptic malignant syndrome; significant risk of suicide or violence; female patients who were pregnant or breast feeding; use of antidepressants or mood stabilizers within 2 weeks of screening; history of any drug sensitivity or hypersensitivity to olanzapine, risperidone, or paliperidone; or a history of unresponsiveness to antipsychotic drugs.

### *Study Conduct*

Researchers at 74 sites with inpatient units in the United States conducted the study.<sup>22</sup> They randomly assigned a total of 444 volunteer patients to one of the 4 “treatment” arms. Previous medications, including antipsychotics, antiparkinsonian agents, and beta-blockers were discontinued 3 days before randomization. Patients were assigned to take either paliperidone 6 mg, paliperidone 12 mg, olanzapine 10 mg, or placebo once daily in the morning for 6 weeks. Participants were hospitalized for the first 14 days and assessed at least weekly throughout the trial. Trained raters assessed drug efficacy using multiple measures, including the PANSS, the Clinical Global Impression (CGI) scale, and the Personal and Social Performance (PSP) scale. The study included multiple safety measures, including laboratory evaluations, physical parameters, and rating scales of neurologic side effects.

The primary efficacy analysis was conducted among randomized patients who received at least one dose of medicine and had at least one post-baseline observation.<sup>22</sup> The Last Observation Carried Forward (LOCF) method was used to deal with missing data for participants who left the study early. Safety was evaluated among all patients who took at least one dose of double-blinded medication.

### *Outcomes*

Of the 444 people randomized, fewer than half of participants in any of the groups completed the 6-week trial (placebo 34%, paliperidone 6 mg 46%, paliperidone 12 mg 48%, and olanzapine 10 mg 45%).<sup>22</sup> In the primary report from this study, the authors presented efficacy data only for paliperidone and placebo. They excluded olanzapine from the primary efficacy analyses because it had been included only for assay sensitivity. Nevertheless, they reported safety outcomes for all treatment arms for the PSP and for many safety measures, including treatment-emergent adverse events.

Using LOCF analyses, the authors reported that both doses of paliperidone were more efficacious than placebo in reducing symptoms measured by the PANSS and that patients taking paliperidone were significantly more likely to meet response criteria.<sup>22</sup> Adverse events rates were high in all groups, including placebo and olanzapine (range 73%–79%). Paliperidone was associated with elevated prolactin levels and dose-related neurologic symptoms, although the specific effects on akathisia and on sexual functioning that might be related to high prolactin levels were not reported. The comparison to olanzapine provided interesting information even though it was included only for assay sensitivity. Overall, olanzapine 10 mg daily appeared to be similar in efficacy to both paliperidone 6 and 12 mg daily, while olanzapine 10 mg seemed to be associated with greater incidence of extrapyramidal side effects than paliperidone 6 mg but lower incidence than paliperidone 12 mg.

### *Limitations of This Efficacy Trial*

Although the low rate of study completion (no study arm with 50% completion rates at 6 weeks) may reflect investigator attitudes about placebo-controlled studies and a desire to minimize the risks to study participants, the huge amount of missing data makes it very difficult to make inferences even about the short-term effects of the drug. An extremely cautious observer might decide the study is inconclusive because less than half of the people who took paliperidone continued to do so for the scheduled 6 weeks of the study. A less conservative interpretation is that paliperidone appeared better than a placebo for reducing schizophrenic symptoms over a few weeks. Assessment of long-term risks and benefits were not goals of this trial.

### **Large Simple Trials: Are There Small but Important Differences Between Available Treatments?**

LSTs focus narrowly on clearly defined, patient-oriented outcomes.<sup>21</sup> A typical LST outcome, mortality, is discrete and meaningful. Designers of LSTs resist the desire to collect information on a wide array of outcomes, instead using resources to enroll large numbers of participants so

they can meaningfully detect relatively small but clinically important differences.<sup>20</sup> LSTs are conducted at typical treatment settings with usual clinical personnel. Study procedures are simple so that the need for specialized research training and interference with routine clinical care is minimized. Inclusion criteria are broad, and exclusion criteria are minimal. The key criterion for study entry is uncertainty about which treatment option is best for the individual participant. Treatments are randomly assigned to avoid selection bias.

#### *Example of a Large Simple Trial*

There are no excellent examples of LSTs involving antipsychotic drugs that have been completed and fully reported. Although the results are not yet available, the best available example is the Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC) trial, which was conducted by Pfizer to satisfy regulatory agencies that wanted to know about the long-term effects of the new drug on mortality.<sup>23</sup> Premarketing studies revealed that ziprasidone was associated with prolongation of the electrocardiographic QT interval, which carried a potential increased risk of fatal ventricular arrhythmias. Because even a small increase in mortality would be clinically important, an LST designed to detect effects on mortality was highly appropriate.

#### *Eligibility*

Individuals aged 18 or older who were diagnosed with schizophrenia and whose psychiatrist wanted to start a new antipsychotic and would consider either olanzapine or ziprasidone were eligible to participate.<sup>23</sup> Exclusion criteria were minimal as follows: pregnant or lactating women, presence of disease with life expectancy less than 1 year, and participation in any study involving investigational products within 30 days of entering the study.

#### *Study Conduct*

Researchers at almost 1000 sites in 18 countries conducted the study.<sup>23</sup> They randomly assigned more than 18 000 volunteer patients to receive open-label olanzapine or ziprasidone. No other study-related interventions, no laboratory testing, or clinical monitoring was required. Minimal data collected at baseline included only the following: demographics, the CGI scale to indicate the severity of illness, cardiac risk factors, and prior antipsychotic use. Simple outcome information, including vital status, data on continuation of the assigned drug, hospitalizations, and emergency room visits, was obtained from the treating physician or other member of the clinical team at regular intervals. If hospitalization occurred, then the hospital records were obtained. Each participant was intended to be followed for 1 year regardless of treatment status. An “Endpoint Committee” that was blinded

to treatment assignments adjudicated the endpoints based on all available information.

#### *Results*

When the results of the ZODIAC study are available, they will provide important information about mortality associated with both ziprasidone and olanzapine over 1 year of treatment.

#### *Limitations of This LST*

Like other LSTs, ZODIAC will provide important data on an important, highly focused research question. The study will provide critical information about the relative risk of death among a broad group of patients treated with either olanzapine or ziprasidone. Because ziprasidone is widely promoted and used among individuals with significant risk factors for cardiovascular disease, its effects on mortality are important. Olanzapine is a reasonable comparator because it too is associated with risk factors for cardiovascular disease, although some observers may have preferred comparing ziprasidone to a drug not so strongly associated with risk factors for heart disease. Secondary outcomes, including hospitalization rates and discontinuation of the assigned study medication, will provide limited information for those who want to know more about the relative effectiveness of these 2 drugs. The ZODIAC study nicely illustrates the trade-offs of a LST—many subjects are enrolled to answer an important, focused question but many other relevant health outcomes are unaddressed. The need to enroll large numbers of participants means that results of LSTs are often not quickly available. For example, the results of the ZODIAC trial, which enrolled more than 18 000 patients and examined relatively long-term outcomes, were still not available several years after the study question was identified and the trial begun.

#### **Practical Trials: How Do Available Treatments Compare Overall in Clinical Practice?**

Practical trials are intended to provide independent evidence to inform decision makers about clinical and policy choices related to the risks and benefits of approved treatments.<sup>24</sup> Researchers design practical trials to provide high-quality evidence, with high internal and external validity, regarding the everyday effectiveness of clinically relevant alternative interventions.<sup>24</sup> To do this, researchers include a heterogeneous population of patients and collect data on a broad range of meaningful health outcomes at many types of practice settings intended to represent usual treatment.<sup>24</sup>

Practical trials use broad subject inclusion criteria and minimal exclusion criteria to enhance external validity and thus enhance the believability of study results for clinicians and patients in typical treatment settings.<sup>25</sup>

Practical trials compare treatments about which there is clinical uncertainty about the outcome at the individual patient level and use randomization to protect against selection biases.<sup>26</sup> Not all practical trials conceal the treatment assignment from patients and study clinicians, but for subjective outcomes determined by raters blinding of the raters is necessary. The training and personnel requirements of rater blinding place burdens on sites that make it less likely that typical clinical sites, rather than research sites, can participate in these trials. In addition, a desire to examine “a broad range of meaningful health outcomes” often including service utilization to allow estimation of cost-effectiveness is at variance with a desire to limit participant and researcher burden. To some extent, practical trials can be conceptualized as hybrids of efficacy and LSTs with the main trade-offs being in internal validity and the potential for a low signal-to-noise ratio.

#### *Example of a Practical (or Pragmatic) Trial*

The US National Institute of Mental Health (NIMH) initiated the CATIE schizophrenia trial to provide objective information about the long-term effectiveness and cost-effectiveness of antipsychotic drugs in widespread use.<sup>10,14</sup> The NIMH supported the study because it had a strong interest in informing policy makers regarding the appropriate use of these commonly used, expensive, and sometimes toxic medications.<sup>27</sup>

The CATIE schizophrenia trial was a hybrid study that included characteristics of efficacy trials and LSTs to examine the overall effectiveness of antipsychotic drugs for chronic schizophrenia.<sup>14</sup> Researchers conducted the study at heterogeneous practice settings, enrolled a relatively large sample of participants, used a discrete primary end point, and attempted to mimic routine clinical practices. However, because CATIE also included double blinding and an extensive array of assessments, many of which required trained clinical raters, all participating clinical sites were required to have specialized research personnel.

#### *Eligibility*

Individuals aged 18–65 years with a diagnosis of schizophrenia and who were appropriate for treatment with an oral antipsychotic medication were eligible to enroll in the CATIE schizophrenia trial.<sup>14</sup> Exclusion criteria were as follows: diagnosis of schizoaffective disorder, mental retardation, or other cognitive disorders; history of serious adverse reactions to any of the proposed treatments; patients having a first schizophrenic episode; patients with persistent severe symptoms despite an adequate trial of one of the proposed treatments, prior treatment with clozapine for treatment resistance; current pregnancy or breast-feeding; and serious and unstable medical conditions. People with medical conditions that were not serious and unstable, and people with substance use disorders, were included.

#### *Study Conduct*

The CATIE schizophrenia trial was conducted at 57 clinical sites representing the array of clinical settings where people with schizophrenia receive treatment in the United States.<sup>10</sup> The study enrolled nearly 1500 participants and randomly assigned them to receive olanzapine, perphenazine, quetiapine, risperidone, or ziprasidone under double-blind conditions.<sup>14</sup> Patients had monthly study visits. Medications were dosed flexibly and, except for additional antipsychotic agents, other medications were permitted throughout the trial. Patients continued on this first treatment for 18 months or until treatment was discontinued for any reason (phase 1).

Participants who did not respond to or tolerate the first assigned medication could enter subsequent phases of the trial. If the phase 1 treatment was perphenazine, patients then received randomized, double-blinded treatment with a second-generation antipsychotic (phase 1B).<sup>14</sup> If patients in phase 1B again discontinued treatment, then they entered phase 2. In phase 2, patients could choose between 2 randomization pathways. An “efficacy” pathway (phase 2E) was recommended to individuals who discontinued the previous treatment due to inadequate symptom relief. A “tolerability” pathway (phase 2T) was recommended to individuals who discontinued the previous treatment due to intolerability. If study participants discontinued either of the phase 2 studies they could enter phase 3, which allowed open choice of one of 9 treatment options.

The CATIE investigators selected time until all-cause treatment discontinuation, meant to reflect the overall effectiveness and acceptability of the medications, as the primary outcome.<sup>14,28</sup> They chose medication discontinuation because it integrates patient and clinician judgments of efficacy, safety, and tolerability into a discrete, global measure of effectiveness. In addition, treatment discontinuation conceptualized as evidence of ineffectiveness avoids the problem of missing data by framing the outcome so that missing data are actually part of the outcome.<sup>21</sup> Secondary outcomes included the reason for treatment discontinuation and changes in PANSS and CGI scores.<sup>14</sup> Safety and tolerability outcomes included adverse events, changes in weight, measures of neurologic side effects, and laboratory analyses of parameters related to lipid and glucose metabolism. Other important outcomes included effects of the drugs on psychosocial and neurocognitive functioning. Health services use was carefully tracked to allow comparison of the cost-effectiveness of the study medicines.

#### *Results*

Almost 1500 individuals with schizophrenia volunteered for the study and took randomly assigned medication.<sup>10</sup> Only a minority of patients in each group remained on the first assigned drug treatment for the 18-month duration of the study (18%–36%). The time to treatment

discontinuation for all causes was longest with olanzapine compared with quetiapine and risperidone. The comparisons of olanzapine with perphenazine and ziprasidone produced similar results but did not reach statistical significance. Substantial increases in weight and adverse metabolic changes associated with increased risk of coronary heart disease were most common with olanzapine.

Analyses of neurocognitive functioning found small but significant improvements for all treatment groups, but no difference among them after 2 months of treatment.<sup>29</sup> Similarly, modest improvements in psychosocial functioning were evident across all treatment groups, but there were no significant differences between treatment groups.<sup>30</sup> Cost-effectiveness analyses also found modest improvement for all groups on several measures of effectiveness over 18 months, but no significant differences between perphenazine and any second-generation antipsychotic.<sup>31</sup> Because total monthly health care costs were lower for perphenazine than for second-generation antipsychotics due to lower antipsychotic drug costs, the investigators concluded that initiating treatment with perphenazine was less costly and no less effective than initiating treatment with each of the 4 newer medications.<sup>31</sup>

In phases 1B, 2E, and 2T, outcomes for individuals taking the various drugs depended considerably upon clinical factors.<sup>11–13</sup> The CATIE researchers concluded that clinical circumstances affect drug effectiveness and that these circumstances should be considered when making drug choices.

### *Limitations of CATIE*

The attempt to satisfy many constituencies by collecting data on many outcomes and to include many treatment arms and randomizations addressing a variety of clinical situations meant that the study design was quite complicated and required trained personnel. Decisions to add ziprasidone to the protocol after recruitment began and to use perphenazine in only one randomized phase of the study generated controversy. The exclusion of people with tardive dyskinesia from possible randomization to perphenazine may have limited the study's ability to detect differences between perphenazine and the newer drugs. The decision to use double-blinded treatments decreased the risk of measurement bias and improved the study's internal validity. However, double blinding increased study costs, decreased the resemblance of the study procedures to those of routine clinical care, and tied the study to relative dose equivalences that have proven controversial. These complexities and its large size made the study expensive, at least compared with other studies not conducted by pharmaceutical companies.

Some aspects of routine clinical practice that were incorporated into the trial design in order to mimic usual

care led to problems interpreting the results and to new information regarding the treatments. For example, because patients entering the study could be randomized to stay on their current treatment and there were different rates of medication use at study entry, the study provided new information about "switching".<sup>32</sup> The researchers found that switching medications affected the primary outcome and that the effect of switching varied according to both the drug being stopped and the one being started.<sup>10,32</sup> Another "naturalistic" aspect of the study, which allowed choice between 2 pathways in phase 2, provided information about patient and clinician preferences but led to small numbers of subjects and low statistical power in the efficacy (clozapine) pathway.

### **Conclusions**

RCTs, with the critical feature of randomization to protect against selection biases, can be the source of high-quality evidence to guide various kinds of health-related decisions. In this article, we discussed some types of late phase trials that can be of use to clinicians and other types of decision makers. We have not included a discussion of all types of RCTs, eg, phase 2 or proof of concept studies, which may focus on dose determination or mechanism of action rather than efficacy.

The goals of a study can help guide decisions regarding appropriate study designs and associated trade-offs. Explanatory or efficacy trials provide ritualized, systematic data regarding the short-term efficacy and safety of new drugs; they provide critical information about how a drug can work in ideal circumstances but provide little information regarding how it might work in usual treatment settings. LSTs can provide definitive answers to a focused clinical question regarding the comparative effects of treatments; the results have tremendous external validity because the studies are conducted in actual treatment settings using routine clinical procedures. If researchers can create effective collaborations with clinicians, these enormous studies can be conducted efficiently and cost-effectively. Although LSTs can answer a focused question definitively, they do not answer the multitude of other health-related questions that might be associated with antipsychotic drugs. Practical trials can provide data regarding many health-related outcomes, but because much data are collected and study logistics are not simple, high costs can limit the numbers of participants studied and may lead to less than definitive results. Practical trials can provide vast amounts of useful information, but designers of these studies must resist the temptation to address all the questions that might be asked regarding a treatment in order to provide optimal information about the most important issues.

The thorough assessment of any health technology, including pharmaceuticals, should include a systematic appraisal of all the available randomized evidence. The confidence that we can have that a drug is likely to be effective in the real world will be increased if a treatment effect found in practical trials is comparable to that observed in highly controlled efficacy trials. For example, the treatment effect on cognition observed in the large pragmatic AD2000 trial comparing donepezil with placebo<sup>33</sup> was comparable to that estimated in the short-term efficacy trials conducted by the manufacturer.<sup>34</sup> However, what can we conclude when trials with different design characteristics differ in their results? We would argue that it cannot be concluded that the results from one kind of trial are necessarily “true” and the results from the other trials are “wrong”. Rather, reasons for the apparent heterogeneity between trials should be sought by examining the clinical and design differences between the trials. If the trials are judged reasonably similar in terms of design and participants, meta-analysis may be possible in which case it is simple enough to determine whether the differences between trials are likely to be real or simply due to the play of chance.

The reasons for any material differences between trials can be assessed using meta-regression, a technique which statistically compares the results of trials according to trial-level characteristics.<sup>35</sup> For example, a systematic review of industry-sponsored trials of second-generation antipsychotics (SGA) compared with first-generation antipsychotics (FGA) found a number of methodological weaknesses in the trials, and, furthermore, a meta-regression strongly suggested that the substantial statistical heterogeneity between trials could be explained by differences in the dose of the FGA used in the trial.<sup>4</sup> The use of a higher dose of a FGA was more likely to produce a result favoring a SGA. Although a subsequent meta-analysis using different methods found partially different results,<sup>36</sup> the design of CATIE was informed by the meta-regression findings on dose. CATIE largely confirmed the conclusion that there were probably no major differences between drugs selected from both available FGAs and SGAs on a broad range of clinically relevant outcomes and cost-effectiveness.<sup>10,29–31</sup> This example demonstrates the particular usefulness of practical trials in situations when there is substantial residual uncertainty about the comparative effectiveness of drugs even in the presence of a large body of efficacy trials. The different objectives of efficacy and practical trials should be considered in the interpretation of the complete body of randomized evidence.

## Disclosure

T.S.S. has consulted or spoken at events sponsored by AstraZeneca, Janssen, Lilly, and Solvay, and he was an investigator in the CATIE trial. JG has received

research funding and support from GlaxoSmithKline and Sanofi-Aventis and has consulted for Bristol-Myers Squibb, TauRx and GlaxoSmithKline. He has received research funding from UK Department of Health, UK Medical Research Council and the Stanley Medical Research Institute.

## References

1. Heres S, Davis J, Maino K, Jetzinger E, Kissling W, Leucht S. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *Am J Psychiatry*. 2006;163:185–194.
2. Freedman R, Carpenter WT Jr, Davis JM, Goldman HH, Tamminga CA, Thomas M. The costs of drugs for schizophrenia. *Am J Psychiatry*. 2006;163:2029–2031.
3. Mintz J, Kopelowicz A. CUtLASS confirms CATIE. *Arch Gen Psychiatry*. 2007;64:978; author reply 979–80.
4. Geddes J, Freemantle N, Harrison P, Bebbington P. Atypical antipsychotics in the treatment of schizophrenia: systematic overview and meta-regression analysis. *BMJ*. 2000;321:1371–6.
5. Lieberman JA, Stroup TS. Guest editors' introduction: what can large pragmatic clinical trials do for public mental health care? *Schizophr Bull*. 2003;29:1–6.
6. Rosenheck RA. Open forum: effectiveness versus efficacy of second-generation antipsychotics: haloperidol without anticholinergics as a comparator. *Psychiatr Serv*. 2005;56:85–92.
7. Chen ZM, Jiang LX, Chen YP, et al. Addition of clopidogrel to aspirin in 45,852 patients with acute myocardial infarction: randomised placebo-controlled trial. *Lancet*. 2005;366:1607–1621.
8. Chen ZM, Pan HC, Chen YP, et al. Early intravenous then oral metoprolol in 45,852 patients with acute myocardial infarction: randomised placebo-controlled trial. *Lancet*. 2005;366:1622–1632.
9. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). *JAMA*. 2002;288:2981–2997.
10. Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med*. 2005;353:1209–1223.
11. McEvoy JP, Lieberman JA, Stroup TS, et al. Effectiveness of clozapine versus olanzapine, quetiapine, and risperidone in patients with chronic schizophrenia who did not respond to prior atypical antipsychotic treatment. *Am J Psychiatry*. 2006;163:600–610.
12. Stroup TS, Lieberman JA, McEvoy JP, et al. Effectiveness of olanzapine, quetiapine, and risperidone in patients with chronic schizophrenia after discontinuing perphenazine: a CATIE study. *Am J Psychiatry*. 2007;164:415–427.
13. Stroup TS, Lieberman JA, McEvoy JP, et al. Effectiveness of olanzapine, quetiapine, risperidone, and ziprasidone in patients with chronic schizophrenia following discontinuation of a previous atypical antipsychotic. *Am J Psychiatry*. 2006;163:611–622.
14. Stroup TS, McEvoy JP, Swartz MS, et al. The national institute of mental health clinical antipsychotic trials of intervention effectiveness (CATIE) project: schizophrenia trial



- design and protocol development. *Schizophr Bull.* 2003;29:15–31.
15. TREC Collaborative Group. Rapid tranquillisation for agitated patients in emergency psychiatric rooms: a randomised trial of midazolam versus haloperidol plus promethazine. *BMJ.* 2003;327:708–713.
  16. Huf G, Coutinho ES, Adams CE. Rapid tranquillisation in psychiatric emergency settings in Brazil: pragmatic randomised controlled trial of intramuscular haloperidol versus intramuscular haloperidol plus promethazine. *BMJ.* 2007;335:869.
  17. Raveendran NS, Tharyan P, Alexander J, Adams CE. Rapid tranquillisation in psychiatric emergency settings in India: pragmatic randomised controlled trial of intramuscular olanzapine versus intramuscular haloperidol plus promethazine. *BMJ.* 2007;335:865.
  18. Jones PB, Barnes TR, Davies L, et al. Randomized controlled trial of the effect on quality of life of second- vs first-generation antipsychotic drugs in schizophrenia: cost utility of the latest antipsychotic drugs in schizophrenia study (CUTLASS 1). *Arch Gen Psychiatry.* 2006;63:1079–1087.
  19. Lewis SW, Barnes TR, Davies L, et al. Randomized controlled trial of effect of prescription of clozapine versus other second-generation antipsychotic drugs in resistant schizophrenia. *Schizophr Bull.* 2006;32:715–723.
  20. Geddes JR. Large simple trials in psychiatry: providing reliable answers to important clinical questions. *Epidemiol Psychiatr Soc.* 2005;14(3):122–126.
  21. Stroup TS, Alves WM, Hamer RM, Lieberman JA. Clinical trials for antipsychotic drugs: design conventions, dilemmas and innovations. *Nat Rev Drug Discov.* 2006;5(2):133–146.
  22. Marder SR, Kramer M, Ford L, et al. Efficacy and safety of paliperidone extended-release tablets: results of a 6-week, randomized, placebo-controlled study. *Biol Psychiatry.* 2007;62(12):134–142.
  23. Strom BL, Faich G, Eng SM, Reynolds RF, Rappard F, Kane JM. Cardiac risk factors and schizophrenia: an analysis of 18,094 patients enrolled in an international comparative trial of olanzapine and ziprasidone. *Eur Psychiatry.* 2007;22:(suppl 1) S111.
  24. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA.* 2003;290:1624–1632.
  25. Stroup S. Practical clinical trials for schizophrenia. *Epidemiol Psychiatr Soc.* 2005;14:132–136.
  26. March JS, Silva SG, Compton S, Shapiro M, Califf R, Krishnan R. The case for practical clinical trials in psychiatry. *Am J Psychiatry.* 2005;162:836–846.
  27. Lebowitz BD, Vitiello B, Norquist GS. Approaches to multi-site clinical trials: the national institute of mental health perspective. *Schizophr Bull.* 2003;29:7–13.
  28. Lieberman JA, Tollefson GD, Charles C, et al. Antipsychotic drug effects on brain morphology in first-episode psychosis. *Arch Gen Psychiatry.* 2005;62:361–370.
  29. Keefe RS, Bilder RM, Davis SM, et al. Neurocognitive effects of antipsychotic medications in patients with chronic schizophrenia in the CATIE trial. *Arch Gen Psychiatry.* 2007;64:633–647.
  30. Swartz MS, Perkins DO, Stroup TS, et al. Effects of antipsychotic medications on psychosocial functioning in patients with chronic schizophrenia: findings from the NIMH CATIE study. *Am J Psychiatry.* 2007;164:428–436.
  31. Rosenheck RA, Leslie DL, Sindelar J, et al. Cost-effectiveness of second-generation antipsychotics and perphenazine in a randomized trial of treatment for chronic schizophrenia. *Am J Psychiatry.* 2006;163:2080–2089.
  32. Essock SM, Covell NH, Davis SM, Stroup TS, Rosenheck RA, Lieberman JA. Effectiveness of switching antipsychotic medications. *Am J Psychiatry.* 2006;163:2090–2095.
  33. Courtney C, Farrell D, Gray R, et al. Long-term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): randomised double-blind trial. *Lancet.* 2004;363:2105–2115.
  34. Whitehead A, Perdomo C, Pratt RD, Birks J, Wilcock GK, Evans JG. Donepezil for the symptomatic treatment of patients with mild to moderate Alzheimer's disease: a meta-analysis of individual patient data from randomised controlled trials. *Int J Geriatr Psychiatry.* 2004;19:624–633.
  35. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med.* 2002;21:1559–1573.
  36. Davis JM, Chen N, Glick ID. A meta-analysis of the efficacy of second-generation antipsychotics. *Arch Gen Psychiatry.* 2003;60:553–564.
  37. Strom BL, Faich G, Eng SM, Reynolds RF, Rappard F, Kane JM. Cardiac risk factors and schizophrenia: an analysis of 18,094 patients enrolled in an international comparative trial of olanzapine and ziprasidone. *Eur. Psychiatry.* 2007;22(Suppl 1):S111.
  38. Stroup TS, McEvoy JP, Swartz MS, et al. The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development. *Schizophr Bull.* 2003;29:15–31.