Celebrating 50 Years!

Evolution of Psychopharmacology
Trial Design and Analysis: Six Decades in the Making

Andrew C. Leon, PhD

Objective: The evolution of trial design and analysis during the lifespan of psychopharmacology is examined.

Background: The clinical trial methodology used to evaluate psychopharmacologic agents has evolved considerably over the past 6 decades. The first and most productive decade was characterized by case series, each with a small number of patients. These trials used nonstandardized clinical observation as outcomes and seldom had a comparison group. The crossover design became widely used to examine acute psychiatric treatments in the 1950s and 1960s. Although this strategy provided comparison data, it introduced problems in study implementation and interpretation. In 1962, the US Food and Drug Administration began to require "substantial evidence of effectiveness from adequate and well-controlled studies." Subsequent decades saw remarkable advances in clinical trial design, assessment, and statistical analyses. Standardized instruments were developed and parallel groups, double-blinding, and placebo controls became the benchmark. Sample sizes increased and data analytic procedures were developed that could accommodate the problems of attrition. Randomized withdrawal designs were introduced in the 1970s to examine maintenance therapies. Ethical principles for research became codified in the United States at that time. A wave of regulatory approvals of novel antipsychotics, antidepressants, and anticonvulsants came in the 1980s and 1990s, each based on data from randomized double-blind, parallel-group, placebo-controlled clinical trials. These trial designs often involved fixed-dose comparisons based, in part, on a greater appreciation that much of the benefit and harm in psychopharmacology was dose related.

Conclusions: Despite the progress in randomized controlled trial (RCT) design, the discovery of new mechanisms of action and blockbuster interventions has slowed during the past decade.

J Clin Psychiatry 2011;72(3):331–340
© Copyright 2011 Physicians Postgraduate Press, Inc.

The development of psychopharmacologic agents over the past 6 decades has been characterized by a paradoxical relationship between medication discovery and clinical trial methodology. The methodology during the most productive decade, 1949–1958, was primitive. Since then, there have been tremendous advances in clinical trial design, assessment, and statistical analyses. Yet, despite numerous innovations in methodology, the discovery of new mechanisms of action and blockbuster interventions seems to have slowed—especially during the past decade. In an effort to understand this phenomenon, the evolution of trial design and analysis during the lifespan of psychopharmacology is examined here.

Initially, the historical context is considered by describing the development of regulatory policy and the early use of clinical trials in medicine. The initial psychopharmacology trials are then reviewed, focusing not on the results, but instead on methodology. Developments over the decades are then examined, culminating with a discussion of the more recent advances in design and analysis. This is not meant to be a comprehensive review of clinical trials in psychopharmacology, but instead a survey of trials that exemplify methodology and its progression over time.

MILESTONES IN US DRUG REGULATION

The US Congress passed the Pure Food and Drugs Act in 1906 to prohibit interstate commerce of misbranded and adulterated foods, drinks, and drugs.[1] The act, motivated in part by problems in the meat packing industry, did not prohibit false therapeutic claims; instead, it focused on ingredients and expanded the authority of the Bureau of Chemistry of the US Department of Agriculture, which was the forerunner of the US Food and Drug Administration (FDA).

There are times that misfortune drives regulatory progress. In 1937, for example, the S. E. Massengill Company of Bristol, Tennessee, prepared a new elixir formulation of sulfanilamide in an effort to provide a palatable alternative to the pill preparation. Tragically, the product contained the solvent diethylene glycol, which killed 107 people, mostly children.[1] This prompted Congress to pass the Federal Food, Drug, and Cosmetic Act in 1938, which required that a manufacturer show that a drug is safe.[1]

In the early 1960s, the sedative and antiemetic thalidomide, which was marketed in Europe, was shown to cause severe birth defects. Frances Kelsey, MD, PhD, a pharmacologist and an FDA medical officer, led efforts to keep thalidomide from the US market. Largely through her efforts the public demanded stronger regulation of drugs. In 1962, Congress passed the Kefauver-Harris Amendments to Federal Food, Drug, and Cosmetic Act, which required that a manufacturer provide *substantial evidence of effectiveness from adequate and well-controlled studies.*[2] In addition, it strengthened drug safety efforts and, most importantly, required that the FDA

approve a drug prior to its marketing.[1] The profound effect of this amendment on psychopharmacology clinical trial methodology will become apparent below.

Experimental Therapeutics

The first randomized controlled clinical trial (RCT) in medicine examined streptomycin for pulmonary tuberculosis and was published in 1948.[3] It applied the randomized study design from agriculture to medical research. It was a randomized controlled, double-blinded clinical trial with 107 participants who were randomly assigned to bed rest either alone or with streptomycin. The 6-month mortality rates were reduced nearly 75%: 27% (bed rest alone) versus 7% (bed rest and medication). Undoubtedly, the difference would have been even much greater had 12- or 24-month mortality been examined.

Initial Psychopharmacology Trials

The initial trials in psychopharmacology involved case series, each with a small number of patients. Cade reported the antimanic properties of lithium based on a series of 10 cases in Australia in 1949.[4] In 1952, the initial psychiatric study of chlorpromazine, which was previously used for nausea in surgical patients, involved 20 patients with psychosis and reported symptomatic improvement.[5] Chlorpromazine was approved by the FDA in 1954 for psychosis. Imipramine has a molecular structure similar to that of chlorpromazine and for that reason was initially tested as an antipsychotic in 1957 with several hundred cases.[6] Although that effort did not demonstrate effectiveness for psychosis, observation of about 12 of the cases with depression revealed the antidepressant property of imipramine. Iproniazid, a monoamine oxidase inhibitor (MAOI), was used for tuberculosis and clinical observation on the tuberculosis wards reported that patients expressed joy and optimism, despite their prognosis. In 1957, a case series of patients with depression showed beneficial effects of iproniazid.[7]

The decade from 1949 to 1958 is unparalleled in the history of psychopharmacology, with the discovery of the first mood stabilizer, the first antipsychotic, and 2 antidepressants, a tricyclic and an MAOI. Yet, none of these case series involved a control. These 4 highly influential case series represent the successes, but do not reveal how many other series showed no effectiveness or revealed safety problems. Further, they do not reveal how many case series were eventually shown to be false positives.

Placebo-Controlled Trials

In 1955, Beecher described placebo response rates across a wide range of indications including anesthesia for surgery, highlighting the need for trials to include a comparator.[8] He stated, "Many a drug has been extolled on the basis of clinical impression when the only power it had was that of a placebo."[8(p1605)] The need for both a control group and double-blinding in experimental research was articulated in 1958.[9]

The first controlled study of lithium involved a placebo-controlled crossover trial.[10] Thirty-eight subjects with mania were enrolled for two 2-week periods. Some cases were open; some blinded. Emotional and motor levels were each rated on a simplistic 3-point scale: +, ++, +++. Among those who were crossed from placebo to lithium, 75% were less manic, whereas none were less manic among those who went from lithium to placebo. Despite this strong evidence, lithium was not approved by the FDA until 1970, due in part to concerns about toxicity.

The first, placebo-controlled trial of chlorpromazine included 12 chronic schizophrenic male inpatients.[11] It was a blinded, crossover study in which subjects were randomized to 1 of 2 sequences with three 6-week periods—chlorpromazine/plcebo/placebo or placebo/placebo/chlorpromazine. No rating scales were used. Based on clinical observation, chlorpromazine significantly reduced "pathological activity." A randomized placebo-controlled trial that specifically recruited subjects with depression showed strong effects of imipramine in 1959.[12]

However, there were many case series results that failed to be confirmed in controlled trials. For instance, 4 case series reported strong antipsychotic properties of reserpine, the Indian herb *Rawoulfia*: 64% marked to moderate improvement,[13] 62%,[14] 46%,[15] and 70%.[16] Yet, none of these had a control. Subsequent controlled trials of reserpine showed no difference from placebo.[17] Another showed no benefit of reserpine, relative to placebo, as an add-on to electroconvulsive therapy.[18]

With the genesis of psychopharmacology, both the National Institute of Mental Health (NIMH) and the Department of Veterans Affairs (VA) set up psychopharmacology research units in the late 1950s. This stimulated the initial stage in the evolution of standards for RCT design and analysis. For instance, Jonathan O. Cole, MD, Director of the NIMH Psychopharmacology Services Center, published recommendations for reporting the results of trials, in which patient selection, evaluation of change, description of treatment setting, and toxicity reactions were all discussed.[19] However, there was no mention of statistics or data analysis.

At the time, study participants were most often inpatients and, when controls were used, crossover designs were the norm. The complexity of the crossover studies escalated. For example, there was a double-blind trial comparing placebo (P), BW203 (B), and chlorpromazine (C).[20] Thirty-six psychotic inpatients were each randomized to 1 of 6 sequences of three 4-week periods: P-B-C, P-C-B, B-P-C, B-C-P, C-P-B, C-B-P. The improvement rates of 62% (placebo), 50% (BW203), and 54% (chlorpromazine) demonstrated that placebo was significantly superior to BW203, undoubtedly the reason that the agent is not familiar to us today.

Inclusion Criteria

The inclusion criteria in the early studies were often rather broad perhaps, in part, because the diagnostic nosology of the era, *DSM-I* (1952) and *DSM-II* (1968), were narrative

based. It was not until Feighner criteria in 1972,[21] Research Diagnostic Criteria in 1978,[22] and *DSM-III* in 1980[23] that nosology became criterion based. In fact, some of the early trials used medication response for diagnostic classification, an approach later referred to as *pharmacologic dissection*.[24] In one such study, 180 subjects with schizophrenia, affective disorders, or other diagnoses showed 7 patterns of response to imipramine including mood elevation, reduction of anxiety, agitated disorganization, and so on.[25] A highly influential example of this approach to diagnoses in clinical trials was a study of imipramine treating 35 inpatients with depression, which found that the recovery rates were markedly different for nondelusional (66.7%) and delusional (23.1%) patients.[26] This study has informed subsequent RCT exclusion criteria and can be thought of as an early example of empirical basis for personalized treatment.

Standards for the study design and analysis continued to evolve. Max Hamilton, MD, a psychiatrist and namesake of a rating scale for depression, published a text that comprised 12 of his lectures covering a range of areas in clinical research design and analysis including stages of experimentation, design of experiments, measurement of variability, tests of statistical significance, *t* test, χ^2, ANOVA, correlation, selecting cases and treatment, and problems in design and analysis.[27]

Innovation in Psychopharmacology Trials

On the heels of the case series and the small RCT paradigm of the 1950s, the scale and complexity changed in the 1960s. Consider, for instance, a VA cooperative study that included 805 subjects with schizophrenia from 37 VA hospitals.[28] It was a double-blinded randomized crossover study with two 12-week periods that compared chlorpromazine, promazine, phenobarbital, and placebo. This study was quite innovative in that it included 2 phenothiazines and an active control (phenobarbital) and used 3 rating scales. The superiority of chlorpromazine was well-documented in this study.

A trial that compared tetrabenazine and chlorpromazine for chronic schizophrenia included 2 novel components: a 6-week washout period and a 2-week placebo lead-in period.[29] However, it did not use randomized treatment assignment; instead, it assigned subjects to the 2 groups (12 weeks of either tetrabenazine or chlorpromazine) matched on age, clinical assessment, behavioral rating, and previous leucotomy. This, like other studies of the time, presented results indicating significant symptomatic improvement *within* each group, but no significant *between* group effect. Such findings highlight the importance of including a comparison group.

After a decade or so of psychopharmacologic research, the standards for design and analysis continued to advance. In a 1962 manuscript on the evaluation of psychopharmacologic agents, Jonathan O. Cole, MD, described methods for each of several diagnostic groups.[30] One area addressed was the conduct of trials for outpatient samples with depression. Several

of the topics covered represent challenges faced in contemporary psychopharmacology: substantial dropout rates with outpatients, the high rate of placebo response, comparative effectiveness, and the response of different subtypes to different agents (ie, personalized treatment). The publication of this comprehensive discussion of clinical trial methodology coincided with the 1962 Kefauver-Harris Amendment to Federal Food, Drug, and Cosmetic Act, which, as stated above, required *substantial evidence of effectiveness from adequate and well-controlled studies*.[2] Although, it is not clear that the publication was motivated by the new legislation, it was at this point in time that the trend in psychopharmacology was shifting from crossover trials to parallel-group designs.

Some studies of that era sought solutions to clinical challenges that we continue to grapple with today. For example, a 9-week placebo-controlled trial of mepazine as an add-on to phenothiazines had cognition as its primary outcome. It did not use clinical observation, but validated scales (Wechsler Adult Intelligence Scale and Hospital Adjustment Scale) to assess outcome.[31] Despite several earlier reports of clearer thinking with mepazine, no group differences were found in this randomized controlled trial.

A landmark study of phenothiazine treatments for acute schizophrenia was conducted during this period.[32] It was a 9-site, randomized, parallel groups, controlled trial that randomized 463 *newly admitted* patients to 6 weeks of chlorpromazine, thioridazine, fluphenazine, or placebo. There were 3 objectives:

(1) Efficacy of thioridazine and fluphenazine relative to placebo.
(2) The noninferiority of thioridazine and fluphenazine to chlorpromazine (although the term *noninferiority* was not used).
(3) Relative safety and tolerability of chlorpromazine, thioridazine, and fluphenazine.

It was this seminal study that developed and first used the now ubiquitous Clinical Global Impressions (CGI)-Severity and CGI-Improvement scales. Participants in this study were terminated due to treatment complications or failures, and the termination rates differed across the cells: active (20%) and placebo (41%). The analyses, which included only the completers, showed significantly greater marked/moderate improvement for the active cells (pooled 75%) versus placebo (23%); however, there were no differences among drugs. In stark contrast to the style used today, the results included the following: "Details on statistical analyses are not reported here. Any differences or relationships reported in this paper, unless otherwise stated, were found to be statistically significant."[32(p252)]

The randomized withdrawal design was introduced into the field of psychopharmacology with 3 trials in the early 1970s. A double-blind lithium discontinuation study in manic-depression ($N = 50$) and recurrent depression

(N = 34) found a significant prophylactic effect of lithium.[33] A small double-blind discontinuation study of lithium in manic-depression and recurrent depression (N = 18) found no difference between lithium and placebo in 2-year relapse rates.[34] A double-blind discontinuation study of recurrent depression compared lithium (N = 22), imipramine (N = 21), and placebo (N = 13) over 2 years.[35] Imipramine had a significantly superior prophylactic effect over placebo.

A NEW WAVE OF DEVELOPMENT IN PSYCHOPHARMACOLOGY

With the 1980s and 1990s came a new wave of regulatory approvals of novel antipsychotics, antidepressants, and anticonvulsants, each based on data from randomized double-blind, parallel-group, placebo-controlled clinical trials. Many of these studies built on the appreciation, developed in the prior decade, that much of the benefit and harm in psychopharmacology was dose related and, therefore, there is a need to apply fixed-dose comparison designs that allow for a brief period of titration. This shift was influenced in part by a letter describing limitations in the interpretation of a therapeutic window for antipsychotics from a flexible-dose study.[36] Dose comparison studies examined haloperidol for acute schizophrenia as well as fluphenazine decanoate[37] and haloperidol decanoate[38] for relapse prevention and provided an opportunity to look closely at extrapyramidal symptoms of haloperidol.[39]

The early fluoxetine studies used dose-escalating schedules in which a dose could range from 20 mg to 80 mg; yet, about 80% of participants got at least 60 mg within 2 weeks.[40,41] However, it was the fixed-dose, dose-response studies that helped identify optimal dosing of fluoxetine with regard to both efficacy and adverse events.[42,43] Furthermore, a study of nonresponders to 3 weeks of 20 mg of fluoxetine compared those who were then randomized to 5 additional weeks of either 20 mg or 60 mg. It found no added benefit of switching to 60 mg and significantly greater attrition due to "adverse experience" for the higher dose.[44]

Identifying the appropriate target population for an intervention is critical. For instance, clozapine was a promising antipsychotic, but the risk of agranulocytosis posed a serious obstacle to regulatory approval. The strategy used to demonstrate the efficacy of clozapine and gain approval in the United States was to conduct a trial in treatment-resistant patients. The pivotal trial recruited participants who had previously failed to respond to at least 3 different neuroleptics.[45] They were initially given 6 weeks lead-in of haloperidol. Only participants who *prospectively* failed to adequately improve during those 6 weeks were then randomized to receive 6 weeks of clozapine or chlorpromazine in a double-blind fashion. Although the response rate was modest for clozapine (30%), it was substantially greater than chlorpromazine (4%).

After a nearly 30-year gap in large-scale double-blind, placebo-controlled drug development employing random assignment to parallel groups for bipolar disorder, maintenance trials of anticonvulsants began a new era. For example, randomized controlled trials compared efficacy and safety of divalproex sodium, lithium carbonate, and placebo[46] and lamotrigine, lithium carbonate, and placebo.[47] Mood stabilizers like valproate typically were shown to have acute efficacy for mania prior to the evaluation of a maintenance effect. Furthermore, the maintenance trials focused on participants recently treated for mania or hypomania.[48] Up until this new era began, the depressed phase of bipolar disorder received considerably less attention, despite its overrepresentation in the course of the illness. Lamotrigine trials provide an exception to this in that the drug was shown to have some evidence of efficacy for acute bipolar depression[49] and subsequently found to provide maintenance therapy for recently depressed participants.[50] As with studies of many psychiatric disorders, trials for bipolar disorder are highly selective, excluding those with psychiatric or other medical comorbidity and alcohol or substance abuse and sometimes those with mixed states or rapid cycling. As a result, the generalizability of results is limited. Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD)[51] and Lithium Treatment–Moderate dose Use Study (LiTMUS) for bipolar disorder[52] each sought to broaden the inclusion criteria.

The field realized the limitations of short-term treatment. Therefore another design was used to investigate treatment during various phases of an illness—sequentially examining acute, continuation, and maintenance phases of treatment. Each phase enrolls successive subsets of participants who met inclusion criteria based on response status in the prior phase. The acute and maintenance phases are each double-blind randomized studies in and of themselves, with randomization at the start of the respective phase. For example, 2 such programs were conducted in chronic depression, one compared sertraline and imipramine[53–55] and the other compared nefazodone and cognitive behavioral analysis system of psychotherapy (CBASP) alone and in combination.[56,57] Each of these programs also included a phase in which acute phase nonresponders were switched to another active agent for acute treatment.[58,59]

The role of psychotherapy augmentation for those not fully responding to an antidepressant was examined in the REVAMP Study. Participants with chronic depression who prospectively failed to respond to algorithm-guided medication were randomized to receive the next level antidepressant either alone or in combination with CBASP or brief supportive psychotherapy.[60]

Although several psychotropic agents have demonstrated efficacy in placebo-controlled trials, there has been limited empirical evidence to guide the choice among efficacious agents for a particular indication. For that reason, among others, the NIMH supported large comparative effectiveness trials including Sequenced Treatment Alternatives to Relieve Depression (STAR*D) and Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) for schizophrenia. These trials involved longer periods of treatment and more

generalizable samples than typically included to date. The CATIE study compared atypical antipsychotics (olanzapine, quetiapine, risperidone, ziprasidone) with the first-generation antipsychotic, perphenazine. The 57-site study enrolled 1,493 participants and used a novel outcome, "time until all cause discontinuation."[61] STAR*D examined treatments for adult outpatients with a nonpsychotic major depressive disorder who did not achieve remission on citalopram therapy. The study used equipoise-stratified randomization[62] in which the participants could opt out of a treatment strategy (switch or augmentation), but not the particular interventions within a strategy.[63] Separate studies examined antidepressant switch strategies (bupropion-SR vs sertraline vs venlafaxine-XR) with 727 participants[64] and antidepressant augmentation strategies (bupropion-SR vs buspirone) with 565 participants.[65]

DESIGNS FOR FUTURE TRIALS IN PSYCHOPHARMACOLOGY

There are 2 promising designs that have been seldom used in psychopharmacology: adaptive design and noninferiority trials. An adaptive design is a "multistage study design that uses accumulating data to decide how to modify aspects of the study without undermining the validity and integrity of the trial."[66(p425)] It is imperative that changes are based on prespecified criteria. For instance, in a dose-finding study, the least effective dose(s) could be dropped after the initial 15% or 20% of planned subjects have completed the study. Alternatively, the randomization allocation ratio might be modified, based on a priori criteria, such that substantially more subjects are randomized to the dose with most promising results to date. Such designs must guard against inflation of type I error and have safeguards that prevent the investigators from learning details of interim results that could have bearing on the remainder of the trial. An independent data monitoring committee might be used to review the interim data and, based on a priori adaptive criteria, convey a general message regarding which changes should be implemented, but not convey the specific results.

Most trials in psychopharmacology use a superiority design, hypothesizing a difference between treatment groups. In contrast, a noninferiority trial is used to show that one cell is *no worse than* the other. It would seem that comparative effectiveness trials could benefit from using the noninferiority design. For example, there would be important policy implications if a trial demonstrated that an inexpensive generic was no worse than a brand name medication. However, there are several fundamental challenges of noninferiority design including demonstration of assay sensitivity, choosing a well-defined margin of noninferiority, and the substantial sample sizes.[67]

Ethical Issues

As a part of the 1962 Kefauver-Harris Amendments, the FDA required that informed consent be obtained from all human research subjects in clinical trials that are submitted

as part of the drug approval process.[1] In 1964, the World Medical Association issued the Declaration of Helsinki that set forth ethical principles for human experimentation. Ethical standards became codified in the United States in the 1970s. The US Congress passed the National Research Act in 1974. This established the commission that issued Belmont Report in 1979,[68] which outlined ethical principles that continue to serve as the basis for the Federal Regulations for protection of human subjects. Despite the standards, ethical perspectives on placebo controls in clinical trials vary from institutional review board (IRB) to IRB and cross-nationally. Policies regarding placebo remain an evolving area in need of harmonization.

IMPACT OF STATISTICAL REASONING ON RESEARCH IN PSYCHOPHARMACOLOGY

The standards for design and analysis were influenced by the initial statisticians involved in psychopharmacology studies. Samuel W. Greenhouse, PhD, was the first statistician at NIMH (1954–1966). C. James Klett, PhD, and John E. Overall, PhD, each played major roles in shaping the quality of the psychopharmacology research of Department of Veterans Affairs Cooperative Studies Program from the late 1950s and beyond. Eugene M. Laska, PhD, joined Rockland State (now the Nathan Kline Institute) as a statistician in 1964. In addition, statisticians were regularly included on NIMH review committees by the 1970s and 1980s; and, in this way, design rigor played a more prominent role in the awarding of research funds. Furthermore, the FDA initiated the multidisciplinary Advisory Committees in 1970s that included biostatisticians.

Data Analyses

The data analytic techniques used in the 1950s and 1960s included χ^2 tests, t tests, analysis of variance, and analysis of covariance. Each of these is useful for comparison of intervention groups in RCTs, yet none adequately accommodates the problem of attrition. The Kaplan-Meier product limit estimate was developed to account for censored cases as a survival analytic approach to cancer research.[69] Due to the influence of Joseph L. Fleiss, PhD, a prominent biostatistician at Columbia University School of Public Health and the New York State Psychiatric Institute, survival analysis was applied to a trial for mania.[70]

Attrition

The initial approach to attrition in psychopharmacology was to limit analyses to participants with complete data. This was a reasonable strategy in the 1950s when studies enrolled only inpatients and dropout was rare, seldom more than 5%, and typically due to death or a rare hospital discharge. Last observation carried forward (LOCF) came into use in the early 1960s, if not before. Another approach involved the replacement of each dropout with a newly randomized subject.

However, the attrition rates became substantially higher over the decades, exceeding 30% in antidepressant trials and 50% in antipsychotic trials.[71] In order to minimize the bias in estimates of the treatment effects in trials with substantial attrition, it is critical to classify participants based on *intention to treat* (ie, randomized assignment), rather than by actual treatment received. This was described by A. Bradford Hill in 1961,[72] yet to this day some investigators resist the proposal to attempt assessing all randomized participants for entire course of RCT, regardless of adherence to study medication, which is arguably the most appropriate implementation of the principle of intention to treat.[73] It was not until the 1980s that statistical strategies accommodated participants with incomplete data.[74,75] Mixed-effects models were introduced in 1982,[74] used in psychopharmacology shortly thereafter,[76–78] but not widely disseminated until the software became accessible, in part with funding from NIMH.[79–82]

The NIMH Treatment of Depression Collaborative Research Program was one of the earliest large trials to apply mixed-effects models, albeit as secondary analyses, in order to include participants with incomplete data.[83] The study randomly assigned 255 subjects to one of four 16-week treatments: cognitive behavior therapy, interpersonal psychotherapy, imipramine hydrochloride plus clinical management, and placebo plus clinical management.[84] It is also noteworthy that this was the first study to develop and incorporate a pharmacotherapy treatment manual to standardize the delivery of a psychopharmacologic intervention in a clinical trial.[85]

Sample Size Determination

Until fairly recently, sample size determination was conducted in rather ad hoc fashion. Even though algorithms and tables for sample size estimates were published in the 1960s,[86,87] sample sizes were typically selected based on 2 criteria: the number of participants included in prior trials and the budget. Power analyses did not become routine until specialized software became available in the 1990s. The need for power analyses for planning clinical trials has now become widely accepted. The concept of the effect size, a fundamental component of power analyses, has gained better understanding. The magnitude of a treatment effect in a completed RCT can be described with an effect size, such as the number needed to treat or area under the curve, each more intuitive than the conventional Cohen *d*.[88] The FDA Division of Psychiatry Products interprets *substantial evidence* primarily as a statistically significant treatment effect. However, a finding that is accompanied by a clinically meaningful effect size carries additional weight.

Assessment

In the 1950s, outcome measures in trials primarily involved clinical observation. There was no standardization across studies. The need for standardized, psychometrically validated assessment tools spurred the development of ratings scales such as the Hamilton Depression Rating Scale,[89] Brief Psychiatric Rating Scale,[90] Montgomery-Asberg Depression Rating Scale,[91] Positive and Negative Syndrome Scale (PANSS),[92] Panic Disorder Severity Scale,[93] Inventory of Depressive Symptomatology,[94] Young Mania Rating Scale,[95] and Clinician-Administered Posttraumatic Stress Disorder Scale.[96] The Early Clinical Drug Evaluation Unit (ECDEU) led an effort to promote uniformity in choice among the many new rating scales by publishing the ECDEU Assessment Manual.[97] More recently the American Psychiatric Association compiled the comprehensive *Handbook of Psychiatric Measures*.[98]

Guidelines for Clinical Trial Design

The momentum for design and analysis standards gained ground, in part, with publications from the NIMH[99] and the American College of Neuropsychopharmacology (ACNP).[100] The FDA published 3 of the initial guidance documents in 1977: General Considerations for the Clinical Evaluation of Drugs,[101] Guidelines for the Clinical Evaluation of Antidepressant Drugs,[102] and Guidelines for the Clinical Evaluation of Antianxiety Drugs.[103] Regulatory guidance continued with the International Conference on Harmonisation (ICH), which published the E9—Statistical Principles for Clinical Trials[104] and E10—Choice of Control Group and Related Issues in Clinical Trials.[105] The FDA continues to develop guidance documents, most recently releasing 2 drafts that are germane to psychopharmacology: Non-Inferiority Clinical Trials[106] and Adaptive Design Clinical Trials for Drugs and Biologics.[107]

A major advance in standardizing content of clinical trial reports came with the introduction of the Consolidated Standards of Reporting Trials (CONSORT).[108] It not only includes the now ubiquitous CONSORT chart showing the flow of participants from screening to study completion, but also presented a 25-item checklist that describes content of various sections of the manuscript including the Title, Abstract, Introduction, Methods, Results, and Discussion. The CONSORT Statement was updated in 2001 and 2010.[109–111]

There had been concern about selective reporting of positive trials and suppression of negative results. This was highlighted, for instance, by the FDA briefing document for the 2004 advisory committee meeting on suicidality and pediatric antidepressant use in which previously unseen negative results were revealed.[112] As part of the 1997 Food and Drug Administration Modernization Act, registration of some clinical trials and presentation of a protocol summary were required in a national database, ClinicalTrials.gov. In 2007, the FDA extended this mandate to include reporting of results and adverse events of completed trials. The International Committee of Medical Journal Editors (ICMJE) initiated a policy that requires investigators to register interventional studies at an acceptable public trials registry (such as ClinicalTrials.gov) as a condition of consideration for publication. It has been a requirement of this journal since 2007.

CONCLUSIONS

Six decades of trials in psychopharmacology were accompanied by major advances in research methodology. The early trials involved a single site, most often an academic medical center that enrolled chronic inpatients who had few, if any, treatment options. Those trials needed a small number of participants to detect the large treatment effects seen with severely ill, treatment naive patients. The patients were very well known to their clinicians, many spending months to years as inpatients in one facility. The clinicians' familiarity with the clinical status of each patient allowed for nonstandardized clinical observations of outcome such as "fewer windows were broken on the ward." The long-term doctor-patient relationships also provided opportunity for serendipity, which formed the foundation for discovery in psychopharmacology in its early decades.

Trial designs evolved from case series with no control to crossover designs to randomized, double-blind, parallel-group placebo-controlled trials. The trials of acute treatment became longer over the decades, initially offering as few as 2 to 4 weeks of treatment and now offering 8 to 12 to 26 weeks. More recently the observed treatment effects have become smaller, requiring multisite, and, more often, multinational and multicontinental studies to provide the number of participants necessary for adequate statistical power.

The paradox that motivated this article was the apparent inconsistency between 6 decades of advances in RCT technology (design, analysis, and assessment) and the slowing of discovery of psychopharmacology. In the decade from 1949 to 1958, 4 major discoveries laid the foundation of psychopharmacology, with lithium, the first mood stabilizer, chlorpromazine, the first antipsychotic, and imipramine and iproniazid, the first antidepressants, each with different mechanisms of action. Why has the discovery of blockbusters slowed today? It could simply stem from retrospective recall bias: were the discovery rates truly much higher in the 1950s? Is this phenomenon a function of publication bias filtering the negative trials or a nostalgic reconstructionist view of the history of psychopharmacology? Are the effect sizes truly shrinking, as has been postulated,[113] or is this phenomenon, in part, a function of trial conduct? Perhaps there is a need for more precise assessment procedures with greater emphasis on reliability and rater training and competence.[114–116] Could it be that today's mental health care delivery system limits the opportunity for serendipitous discovery, a driving force in early psychopharmacology? In my interviews of several who helped shape the field, I was repeatedly told that an insufficient amount of time is spent in phase 2 development to determine the correct drug, the proper dose, and the appropriate patient population. This suggests that the hurried effort to advance the development and regulatory approval of psychopharmacologic compounds could, in fact, have set the stage to miss potential blockbusters that were inadequately tested.

An immediate challenge faced in the field is to make progress in the development and identification of personalized treatments,[117] perhaps through the application of biomarkers. The concept of identifying moderators of the between-treatment effect size was articulated for clinical trials in psychiatry[118] and will no doubt be applied in the effort to uncover personalized treatments.

Author affiliation: Weill Cornell Medical College, New York, New York.

REFERENCES

1. US Food and Drug Administration. US Food and Drug Administration: History. 2010. http://www.fda.gov/AboutFDA/WhatWeDo/History/FOrgsHistory/CDER/CenterforDrugEvaluationandResearchBrochureandChronology/ucm114470.htm#1906. Accessed November 23, 2010.
2. US Food and Drug Administration. *Federal Food, Drug, and Cosmetic, As Amended, FDA 93-1051*. Washington, DC: Superintendent of Documents, US Government Printing Office; 1993.
3. Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *BMJ*. 1948;2(4582):769–782.
4. Cade JFJ. Lithium salts in the treatment of psychotic excitement. *Med J Aust*. 1949;2(10):518–520.
5. Delay J, Deniker P, Harl JM. Utilisation en therapeutique psychiatrique dune phenothiazine daction centrale elective (4560-Rp). *Presse Med*. 1952;60(64):1369.
6. Kuhn V. Uber die Behandlung depressiver zustande mit einem iminodibenzylderival (G 22355). *Schweiz Med Wochenschr*. 1957;(35–36):1135–1140.
7. Loomer HP, Saunders JC, Kline NS. A clinical and pharmacodynamic evaluation of iproniazid as a psychic energizer. *Psychiatr Res Rep Am Psychiatr Assoc*. 1957;8:129–141.
8. Beecher HK. The powerful placebo. *J Am Med Assoc*. 1955;159(17):1602–1606.
9. Modell W, Houde RW. Factors influencing clinical evaluation of drugs; with special reference to the double-blind technique. *J Am Med Assoc*. 1958;167(18):2190–2199.
10. Schou M, Juel-Nielsen N, Stromgren E, et al. The treatment of manic psychoses by the administration of lithium salts. *J Neurol Neurosurg Psychiatry*. 1954;17(4):250–260.
11. Cutler RP, Monroe JJ, Anderson TE. Effects of tranquilizers upon pathological activity in psychotic patients. *AMA Arch Neurol Psychiatry*. 1957;78(1):61–68.

12. Ball JR, Kiloh LG. A controlled trial of imipramine in treatment of depressive states. *BMJ*. 1959;2(5159):1052–1055.

13. Noce RH, Williams DB, Rapaport W. Reserpine (Serpasil) in the management of the mentally ill. *J Am Med Assoc*. 1955;158(1):11–15.

14. Barsa JA, Kline NS. Combined reserpine-chlorpromazine therapy in disturbed psychotics. *Am J Psychiatry*. 1955;111(10):780.

15. Hollister LE, Krieger GE, Kringel A, et al. Treatment of chronic schizophrenic reactions with reserpine. *Ann N Y Acad Sci*. 1955;61(1):92–100.

16. Kinross-Wright V. Chlorpromazine and reserpine in the treatment of psychoses. *Ann N Y Acad Sci*. 1955;61(1):174–182.

17. Campden-Main BC, Wegielski Z. The control of deviant behavior in chronically disturbed psychotic patients by the oral administration of reserpine. *Ann N Y Acad Sci*. 1955;61(1):117–122.

18. Goller ES. A controlled trial of reserpine in chronic schizophrenia. *J Ment Sci*. 1960;106:1408–1412.

19. Cole JO. Recommendations for reporting studies of psychiatric drugs. *Public Health Rep*. 1957;72(7):638–645.

20. Fleming BG, Currie JD. Investigation of a new compound, BW203, and of chlorpromazine in the treatment of psychosis. *J Ment Sci*. 1958;104(436):749–757.

21. Feighner JP, Robins E, Guze SB, et al. Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatry*. 1972;26(1):57–63.

22. Spitzer RL, Endicott J, Robins E. Research Diagnostic Criteria: rationale and reliability. *Arch Gen Psychiatry*. 1978;35(6):773–782.

23. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, Third Edition. Washington, DC: American Psychiatric Association; 1980.

24. Quitkin FM, McGrath PJ, Stewart JW, et al. Phenelzine and imipramine in mood reactive depressives: further delineation of the syndrome of atypical depression. *Arch Gen Psychiatry*. 1989;46(9):787–793.

25. Klein DF, Fink M. Psychiatric reaction patterns to imipramine. *Am J Psychiatry*. 1962;119:432–438.

26. Glassman AH, Kantor SJ, Shostak M. Depression, delusions, and drug response. *Am J Psychiatry*. 1975;132(7):716–719.

27. Hamilton M. *Lectures on the Methodology of Clinical Research*. Edinburgh, Scotland: E&S Livingstone; Williams & Wilkins; 1961.

28. Casey JF, Bennett IF, Lindley CJ, et al. Drug therapy in schizophrenia: a controlled study of the relative effectiveness of chlorpromazine, promazine, phenobarbital, and placebo. *Arch Gen Psychiatry*. 1960;2:210–220.

29. Ashcroft GW, MacDougall EJ, Barker PA. A comparison of tetrabenazine and chlorpromazine in chronic schizophrenia. *J Ment Sci*. 1961;107:287–293.

30. Cole JO. Evaluation of drug treatments in psychiatry. *J New Drugs*. 1962;2:264–275.

31. Whittier JR, Klein DF, Levine G, et al. Mepazine (Pacatal): clinical trial with placebo control and psychological study. *Psychopharmacologia*. 1960;1(4):280–287.

32. National Institute of Mental Health Psychopharmacology Service Center Collaborative Study Group. Phenothiazine treatment in acute schizophrenia; effectiveness. *Arch Gen Psychiatry*. 1964;10:246–261.

33. Baastrup PC, Poulsen JC, Schou M, et al. Prophylactic lithium: double blind discontinuation in manic-depressive and recurrent-depressive disorders. *Lancet*. 1970;2(7668):326–330.

34. Melia PI. Prophylactic lithium: a double-blind trial in recurrent affective disorders. *Br J Psychiatry*. 1970;116(535):621–624.

35. Prien RF, Caffey EM Jr, Klett CJ. Prophylactic efficacy of lithium carbonate in manic-depressive illness: report of the Veterans Administration and National Institute of Mental Health collaborative study group. *Arch Gen Psychiatry*. 1973;28(3):337–341.

36. Van Putten T, Marder SR. Variable dose studies provide misleading therapeutic windows. *J Clin Psychopharmacol*. 1986;6(4):249–250.

37. Marder SR, Van Putten T, Mintz J, et al. Low- and conventional-dose maintenance therapy with fluphenazine decanoate: two-year outcome. *Arch Gen Psychiatry*. 1987;44(6):518–521.

38. Kane JM, Davis JM, Schooler N, et al. A multidose study of haloperidol decanoate in the maintenance treatment of schizophrenia. *Am J Psychiatry*. 2002;159(4):554–560.

39. Van Putten T, Marder SR, Mintz J. A controlled dose comparison of haloperidol in newly admitted schizophrenic patients. *Arch Gen Psychiatry*. 1990;47(8):754–758.

40. Chouinard G. A double-blind controlled clinical trial of fluoxetine and amitriptyline in the treatment of outpatients with major depressive disorder. *J Clin Psychiatry*. 1985;46(3 Pt 2):32–37.

41. Feighner JP, Cohn JB. Double-blind comparative trials of fluoxetine and doxepin in geriatric patients with major depressive disorder. *J Clin Psychiatry*. 1985;46(3 Pt 2):20–25.

42. Fabre LF, Putman HP 3rd. A fixed-dose clinical trial of fluoxetine in outpatients with major depression. *J Clin Psychiatry*. 1987;48(10):406–408.

43. Wernicke JF, Dunlop SR, Dornseif BE, et al. Fixed-dose fluoxetine therapy for depression. *Psychopharmacol Bull*. 1987;23(1):164–168.

44. Dornseif BE, Dunlop SR, Potvin JH, et al. Effect of dose escalation after low-dose fluoxetine therapy. *Psychopharmacol Bull*. 1989;25(1):71–79.

45. Kane J, Honigfeld G, Singer J, et al. Clozapine for the treatment-resistant schizophrenic: a double-blind comparison with chlorpromazine. *Arch Gen Psychiatry*. 1988;45(9):789–796.

46. Bowden CL, Calabrese JR, McElroy SL, et al; Divalproex Maintenance Study Group. A randomized, placebo-controlled 12-month trial of divalproex and lithium in treatment of outpatients with bipolar I disorder. *Arch Gen Psychiatry*. 2000;57(5):481–489.

47. Bowden CL, Calabrese JR, Sachs G, et al; Lamictal 606 Study Group. A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently manic or hypomanic patients with bipolar I disorder. *Arch Gen Psychiatry*. 2003;60(4):392–400.

48. Bowden CL, Brugger AM, Swann AC, et al; The Depakote Mania Study Group. Efficacy of divalproex vs lithium and placebo in the treatment of mania. *JAMA*. 1994;271(12):918–924.

49. Calabrese JR, Bowden CL, Sachs GS, et al. A double-blind placebo-controlled study of lamotrigine monotherapy in outpatients with bipolar I depression: Lamictal 602 Study Group. *J Clin Psychiatry*. 1999;60(2):79–88.

50. Calabrese JR, Bowden CL, Sachs G, et al; Lamictal 605 Study Group. A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently depressed patients with bipolar I disorder. *J Clin Psychiatry*. 2003;64(9):1013–1024.

51. Sachs GS, Thase ME, Otto MW, et al. Rationale, design, and methods of the systematic treatment enhancement program for bipolar disorder (STEP-BD). *Biol Psychiatry*. 2003;53(11):1028–1042.

52. Nierenberg AA, Sylvia LG, Leon AC, et al; Litmus Study Group. Lithium treatment: moderate dose use study (LiTMUS) for bipolar disorder: rationale and design. *Clin Trials*. 2009;6(6):637–648.

53. Rush AJ, Koran LM, Keller MB, et al. The treatment of chronic depression, part 1: study design and rationale for evaluating the comparative efficacy of sertraline and imipramine as acute, crossover, continuation, and maintenance phase therapies. *J Clin Psychiatry*. 1998;59(11):589–597.

54. Keller MB, Gelenberg AJ, Hirschfeld RM, et al. The treatment of chronic depression, part 2: a double-blind, randomized trial of sertraline and imipramine. *J Clin Psychiatry*. 1998;59(11):598–607.

55. Keller MB, Kocsis JH, Thase ME, et al. Maintenance phase efficacy of sertraline for chronic depression: a randomized controlled trial. *JAMA*. 1998;280(19):1665–1672.

56. Keller MB, McCullough JP, Klein DN, et al. A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *N Engl J Med*. 2000;342(20):1462–1470.

57. Klein DN, Santiago NJ, Vivian D, et al. Cognitive-behavioral analysis system of psychotherapy as a maintenance treatment for chronic depression. *J Consult Clin Psychol*. 2004;72(4):681–688.

58. Thase ME, Rush AJ, Howland RH, et al. Double-blind switch study of imipramine or sertraline treatment of antidepressant-resistant chronic depression. *Arch Gen Psychiatry*. 2002;59(3):233–239.

59. Schatzberg AF, Rush AJ, Arnow BA, et al. Chronic depression: medication (nefazodone) or psychotherapy (CBASP) is effective when the other is not. *Arch Gen Psychiatry*. 2005;62(5):513–520.

60. Kocsis JH, Gelenberg AJ, Rothbaum BO, et al; REVAMP Investigators. Cognitive behavioral analysis system of psychotherapy and brief supportive psychotherapy for augmentation of antidepressant nonresponse in chronic depression: the REVAMP Trial. *Arch Gen Psychiatry*. 2009;66(11):1178–1188.

61. Lieberman JA, Stroup TS, McEvoy JP, et al; Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Investigators. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med*. 2005;353(12):1209–1223.

62. Lavori PW, Rush AJ, Wisniewski SR, et al. Strengthening clinical effectiveness trials: equipoise-stratified randomization. *Biol Psychiatry*. 2001;50(10):792–801.

63. Rush AJ, Fava M, Wisniewski SR, et al; STAR*D Investigators Group. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials*. 2004;25(1):119–142.

64. Rush AJ, Trivedi MH, Wisniewski SR, et al; STAR*D Study Team. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs

for depression. *N Engl J Med*. 2006;354(12):1231–1242.

65. Trivedi MH, Fava M, Wisniewski SR, et al; STAR*D Study Team. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med*. 2006;354(12):1243–1252.

66. Dragalin V. Adaptive designs: terminology and classification. *Drug Inf J*. 2006;40:425–435.

67. Leon AC. Comparative effectiveness clinical trials in psychiatry: superiority, noninferiority and the role of active comparators [published online ahead of print February 8, 2011]. *J Clin Psychiatry*.

68. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont report: ethical principles and guidelines for the protection of human subjects of research*. http://ohsr.od.nih.gov/guidelines/belmont.html. Updated April 18, 1979. Accessed December 20, 2010.

69. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457–481.

70. Fleiss JL, Dunner DL, Stallone F, et al. The life table: a method for analyzing longitudinal studies. *Arch Gen Psychiatry*. 1976;33(1):107–112.

71. Leon AC, Mallinckrodt CH, Chuang-Stein C, et al. Attrition in randomized controlled clinical trials: methodological issues in psychopharmacology. *Biol Psychiatry*. 2006;59(11):1001–1005.

72. Hill AB. *Principles of Medical Statistics*. 7th ed. New York, NY: Oxford University Press; 1961.

73. Lavori PW. Clinical trials in psychiatry: should protocol deviation censor patient data? *Neuropsychopharmacology*. 1992;6(1):39–48, discussion 49–63.

74. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982;38(4):963–974.

75. Jennrich RI, Schluchter MD. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*. 1986;42(4):805–820.

76. Gibbons RD, Hedeker D, Waternaux C, et al. Random regression models: a comprehensive approach to the analysis of longitudinal psychiatric data. *Psychopharmacol Bull*. 1988;24(3):438–443.

77. Hedeker D, Gibbons RD, Waternaux C, et al. Investigating drug plasma levels and clinical response using random regression models. *Psychopharmacol Bull*. 1989;25(2):227–231.

78. Hedeker D, Gibbons RD, Davis JM. Random regression models for multicenter clinical trials data. *Psychopharmacol Bull*. 1991;27(1):73–77.

79. Hedeker D. MIXOR: a Fortran program for mixed-effects ordinal probit and logistic regression. *Technical Report, Prevention Research Center*. Chicago, IL: School of Public Health, University of Illinois at Chicago; 1992.

80. Hedeker D. MIXREG: a Fortran program for mixed-effects linear regression with autocorrelated errors. *Technical Report, Prevention Research Center*. Chicago, IL: School of Public Health, University of Illinois at Chicago; 1992.

81. Hedeker D, Gibbons RD. MIXREG: a computer program for mixed-effects regression analysis with autocorrelated errors. *Comput Methods Programs Biomed*. 1996;49(3):229–252.

82. Hedeker D, Gibbons RD. MIXOR: a computer program for mixed-effects ordinal regression analysis. *Comput Methods Programs Biomed*. 1996;49(2):157–176.

83. Gibbons RD, Hedeker D, Elkin I, et al. Some conceptual and statistical issues in analysis of longitudinal psychiatric data: application to the NIMH treatment of Depression Collaborative Research Program dataset. *Arch Gen Psychiatry*. 1993;50(9):739–750.

84. Elkin I, Shea MT, Watkins JT, et al. National Institute of Mental Health Treatment of Depression Collaborative Research Program: general effectiveness of treatments. *Arch Gen Psychiatry*. 1989;46(11):971–982, discussion 983.

85. Fawcett J, Epstein P, Fiester SJ, et al; NIMH Treatment of Depression Collaborative Research Program. Clinical management—imipramine/placebo administration manual. *Psychopharmacol Bull*. 1987;23(2):309–324.

86. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Academic Press; 1969.

87. Overall JE, Dalal SN. Design of experiments to maximize power relative to cost. *Psychol Bull*. 1965;64(5):339–350.

88. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry*. 2006;59(11):990–996.

89. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23(1):56–62.

90. Overall JE. The Brief Psychiatric Rating Scale. *Psychopharmacol Bull*. 1988;24:97–99.

91. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134(4):382–389.

92. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13(2):261–276.

93. Shear MK, Brown TA, Barlow DH, et al. Multicenter collaborative Panic Disorder Severity Scale. *Am J Psychiatry*. 1997;154(11):1571–1575.

94. Rush AJ, Gullion CM, Basco MR, et al. The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychol Med*. 1996;26(3):477–486.

95. Young RC, Biggs JT, Ziegler VE, et al. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry*. 1978;133(5):429–435.

96. Blake DD, Weathers FW, Nagy LM, et al. The development of a Clinician-Administered PTSD Scale. *J Trauma Stress*. 1995;8(1):75–90.

97. Guy W. *ECDEU Assessment Manual For Psychopharmacology*. Rockville, MD: Department of Health, Education, and Welfare; 1976.

98. Rush AJ, First MB, Blacker D, American Psychiatric Association. *Task Force for the Handbook of Psychiatric Measures. Handbook of Psychiatric Measures*. 2nd ed. Washington, DC: American Psychiatric Publishing.; 2008.

99. Levine J, Schiele BC, Bouthilet L. *Principles and Problems in Establishing the Efficacy of Psychotropic Agents*. Chevy Chase, MD: National Institute of Mental Health; 1971.

100. Prien RF, Robinson DS, National Institute of Mental Health, American College of Neuropsychopharmacology. *Clinical Evaluation of Psychotropic Drugs: Principles and Guidelines*. New York, NY: Raven Press; 1994.

101. Food and Drug Administration. General Considerations for the Clinical Evaluation of Drugs. 1977. http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/WomensHealthResearch/UCM131196.pdf. Accessed January 28, 2011.

102. Food and Drug Administration. Guidelines for the Clinical Evaluation of Antidepressant Drugs; 1977. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071299.pdf. Updated February 1997. Accessed December 20, 2010.

103. Food and Drug Administration. Guidelines for the Clinical Evaluation of Antianxiety Drugs 1977. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071295.pdf. Updated February 1997. Accessed December 20, 2010.

104. Food and Drug Administration Conference on Harmonisation I. E9—Statistical Principles for Clinical Trials 1998. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073137.pdf. Updated September 1998. Accessed December 20, 2010.

105. Food and Drug Administration Conference on Harmonization I. E10—Choice of Control Group and Related Issues in Clinical Trials 2001. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073139.pdf. Updated May 2001. Accessed December 20, 2010.

106. Food and Drug Administration. Non-Inferiority Clinical Trials: Draft Guidance for Industry; 2010. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf. Updated March 2010. Accessed December 20, 2010.

107. Food and Drug Administration. Adaptive Design Clinical Trials for Drugs and Biologics: Draft Guidance for Industry; 2010. http://www.fda.gov/downloads/Drugs/guidancecomplianceregulatoryinformation/guidances/ucm201790.pdf. Updated February 2010. Accessed December 20, 2010.

108. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA*. 1996;276(8):637–639.

109. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001;357(9263):1191–1194.

110. Schulz KF, Altman DG, Moher DCONSORT; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med*. 2010;7(3):e1000251.

111. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.

112. Food and Drug Administration. (2004). Briefing document for February 2, 2004 Meeting of Psychopharmacological Drugs Advisory Committee (PDAC) and Pediatric Subcommittee of the Anti-Infective Drugs Advisory Committee (Peds AC). http://www.fda.gov/ohrms/dockets/ac/04/briefing/2004-4065b1-04-Tab02-Laughren-Jan5.pdf. Accessed January 28, 2011.

113. Walsh BT, Seidman SN, Sysko R, et al. Placebo response in studies of major depression: variable, substantial, and growing.

JAMA. 2002;287(14):1840–1847.

114. Kraemer HC. To increase power in randomized clinical trials without increasing sample size. *Psychopharmacol Bull.* 1991;27(3): 217–224.

115. Leon AC, Marzuk PM, Portera L. More reliable outcome measures can reduce sample size requirements. *Arch Gen Psychiatry.* 1995; 52(10):867–871.

116. Kobak KA, Kane JM, Thase ME, et al. Why do clinical trials fail? the problem of measurement error in clinical trials: time to test new paradigms? *J Clin Psychopharmacol.* 2007;27(1):1–5.

117. Leon AC. Two clinical trial designs to examine personalized treatments for psychiatric disorders [article published online ahead of print July 13, 2010]. *J Clin Psychiatry.* 2010.

118. Kraemer HC, Wilson GT, Fairburn CG, et al. Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry.* 2002;59(10):877–883.