

Evidence from Clinical Trials: Can We Do Better?

Andrew D. Siderowf

Department of Neurology, University of Pennsylvania, Philadelphia, Pennsylvania 19107

Summary: Randomized clinical trials provide the most internally valid evidence for medical decision-making. In many areas of neurology, results from clinical trials showing which therapies are and are not effective have had a substantial impact on patient care. Relative to observational methods, the central advantage of clinical trials is control of bias attributable to unmeasured differences between patients. However, trials also have clear limitations, including a historical failure to include a representative cross-section of patients with a given disease,

and highly structured treatment regimes that are difficult to replicate in normal practice settings. These limitations tend to reduce the generalizability of results from clinical trials. This article reviews some ways in which the design and application of clinical trials could be improved so that the evidence produced would be more relevant to health-care providers and other decision makers. **Key Words:** Clinical trials, neurology, health measurement, placebo, randomized, bias.

INTRODUCTION

In a recent attempted meta-analysis, Smith and Pell¹ reviewed the available randomized trials supporting the use of parachutes to prevent injuries caused by jumping out of an airplane. These authors found that no trials had been done and concluded that there was insufficient evidence to recommend the use of parachutes. This paper reminds us that some interventions are of such intuitive value that they do not require randomized clinical trials. However, very few interventions in medicine work quite as definitively as parachutes. For the rest, we require evidence to convince us of their merit. The most reliable evidence in medicine comes from blinded, randomized, controlled trials. It is the only design paradigm that reliably controls for unobserved differences between treated and untreated patients. In the recent past, clinical trials have been conducted that have advanced the understanding of neurological disease and improved the care of patients with neurological disorders. However, there is clearly room for improvement. Much work remains to make trials more relevant to the decision-making of patients, physicians, and health policy makers.

ADVANTAGES OF CLINICAL TRIALS OVER OBSERVATIONAL DESIGNS

Because clinical trials are scientific experiments on human subjects, they should only be performed when the knowledge to be gained from the trial is important, and when the information cannot be obtained through other research designs. The most rigorous trials employ both random assignment of patients and blinding of subjects and investigators to treatment assignment.² Randomization ensures that factors that may be related to the likelihood of experiencing an outcome are evenly distributed among subjects in each treatment group (selection bias). Blinding prevents investigators or patients from applying different standards to their assessments of treatment effect based on knowledge of which treatment they are receiving (information bias). Although both randomization and blinding are important contributors to the quality of a clinical trial, randomization may be more important,² and blinding may be difficult to achieve in some settings. The North American Symptomatic Carotid Endarterectomy Trial³ and the Stroke Prevention in Atrial Fibrillation III trial⁴ are examples of trials that employed randomization but not blinding, and produced valid and influential results. Unlike observational studies, only randomized trials reliably control for unidentifiable differences between subjects and provide unbiased estimates of the effects of treatment.⁵ Even when the treatment effects seen in open studies appear unmistakable, it is quite possible for uncontrolled studies to be followed by

Address correspondence and reprint requests to Andrew Siderowf, M.D., Department of Neurology, 330 South Ninth Street, Second Floor, Philadelphia, PA 19107. E-mail: adsiderowf@pahosp.com.

randomized, controlled studies showing no effect, or even harmful effects from treatment.

The case of hormone replacement therapy to prevent atherosclerotic vascular disease is the most commonly cited example in which observational evidence showing an advantage for a therapy was subsequently questioned by a randomized trial.^{6,7} In retrospect, it is possible to see that subtle differences between subjects who were and were not treated on a number of healthy lifestyle practices probably accounted for the effects seen in the observational studies. Following the reporting of the randomized trial, the use of hormone replacement therapy in postmenopausal women has fallen dramatically.⁸

There are also examples from neurology of therapies that appeared extremely promising based on uncontrolled trials, but were not efficacious in blinded, controlled trials. Fetal cell transplantation for Parkinson's disease (PD) is an example of this problem. The promising results from observational studies^{9,10} were not supported by subsequent randomized, controlled trials.^{11,12} The results of these trials have led investigators to return to preclinical studies to understand how to translate this promising technology into a viable therapy. The use of heparinoids in acute stroke is another example. Observational studies¹³ and expert opinion¹⁴ had supported the use of heparinoids for patients with recent cerebral ischemia. However, large-scale randomized trials such as the International Stroke Trial¹⁵ showed that the small benefits of heparin for reducing recurrent ischemia are offset by increased risk of bleeding complications.

Possibly the major weakness of observational studies that compare one therapeutic strategy with another is that treating physicians may select patients for a given therapy based on clinical features that are also related to the outcomes of interest. This type of confounding is known as "confounding-by-indication."¹⁶ If it were possible to completely control for the factors upon which patients were selected, confounding-by-indication might not be a problem. However, because there are often unmeasured as well as measured factors related to selection of patients for a given treatment, confounding-by-indication may be accompanied by bias. It is this associated bias that threatens the validity of observational studies. The perspective that the randomized, controlled trial is the most reliable type of medical "evidence" has been reinforced by evidence rating systems¹⁷ that strongly favor clinical trials relative to observational methods.

EXAMPLES OF TRIALS THAT HAVE IMPACTED TREATMENT OF NEUROLOGICAL DISEASE

One need not search for very long to identify recent examples of clinical trials that have contributed to various areas of neurology. Several of these studies have

been conducted in the area of stroke or cerebrovascular disease. The International Stroke Trial¹⁵ randomized 19,435 patients around the world, showed only slight treatment effects in favor of heparin, and showed an increased risk of bleeding at higher doses. The Warfarin and Aspirin for Recurrent Stroke (WARSS) trial¹⁸ compared the effectiveness of aspirin and warfarin for preventing recurrent stroke and found no benefit for warfarin. These results have provided important evidence to address the potential overuse of harmful anticoagulation that had been considered reasonable practice for many years. On the other hand, the trial of intravenous recombinant tissue plasminogen activator, sponsored by the National Institute of Neurological Disorders and Stroke,¹⁹ provided the first evidence of an effective thrombolytic treatment for acute stroke, assuming this therapy is given within 3 h of symptom onset, and the Chinese Acute Stroke Trial randomized a similar number of subjects and found a benefit for aspirin treatment within the first 48 h after symptom onset on stroke outcomes.²⁰

In epilepsy and in Parkinson's disease there have been randomized, controlled trials of surgical interventions that have provided highly informative results. Weibe and colleagues²¹ compared temporal lobe epilepsy surgery to medical management for patients with intractable complex partial epilepsy and showed the benefits of surgical intervention. By contrast, two sets of investigators^{11,12} found minimal clinical benefits, and possibly substantial harm from intrastriatal implantation of fetal dopaminergic cells in patients with advanced Parkinson disease. Although these surgical trials were "positive" in one clinical setting and "negative" in the other, they both illustrate the importance of conducting rigorous evaluations of surgical as well as medical therapies.

In the area of demyelinating disease, a number of trials have ushered in a new era of immunomodulatory therapy.²²⁻²⁴ These trials have demonstrated the potential to modify the long-term course of a chronic neurological disease. They have also demonstrated the potential to use imaging biomarkers as a complementary endpoint in a neurological clinical trial.²⁵ The Optic Neuritis Treatment Trial²⁶ and Controlled High-Risk Subjects Avonex Multiple Sclerosis Prevention Study trial²³ suggest that it may be possible to intervene to prevent the development of chronic multiple sclerosis (MS) after the first demyelinating episode. These examples, as well as those from other neurological areas cited above, demonstrate that recent clinical trials have provided reliable evidence that has impacted the care of patients with neurological disorders.

IMPROVING TRIALS: WE CAN DO BETTER

There have clearly been successes among trials for neurological disease, but much room for improvement

remains. There are at least four general areas in which trial methods could be improved to increase the clinical relevance and overall usefulness of trial results. These areas include the following: 1) increasing the applicability of trial results by including subjects from traditionally underrepresented sociodemographic groups and employing more naturalistic approaches to treatment, 2) conducting more comparative trials, in which the hypothesis is specifically intended to address a treatment decision rather than proof of biological activity, 3) expanding the uses of trials to address interventions beyond chemotherapeutics, and 4) improving methods for investigating the long-term effects of treatment for chronic neurological conditions. In several of these areas, improvements are underway, and progress has already been made toward providing physicians and other decision-makers with the information they need to provide better care for patients with neurological disorders.

The lack of generalizability (or external validity) is usually considered the major limitation of randomized trials. Formally, generalizability can be defined as the extent to which the results of a trial provide a correct basis for generalizations to other circumstances.²⁷ Patients who participate in clinical trials may be selected based on factors that the investigator believes make them good study subjects, and they may not necessarily be representative of the overall population of patients with a given disease. For example, patients that have been included in reported case series describing the effectiveness of deep brain stimulation for PD are younger than most patients with advanced, medically refractory disease who would be considered candidates for deep brain stimulation. In this case, the results of the study must be extrapolated to patient groups, such as older patients, that may benefit from treatment but were not explicitly included in the trial. Likewise, the treatment milieu of a trial may not be easily reproduced outside of specialized centers. Therefore, the results of the trial may not be applicable to some treatment settings, even though the patients there have similar signs and symptoms. Phase IV, or postmarketing studies which study the “real-life” application of new treatments can be useful in determining how such treatments work outside of the clinical trial setting.

Expanding on the idea of generalizability, Dans and colleagues²⁸ have put forward the concept of applicability in deciding whether the results of a clinical trial are likely to be relevant to a specific patient. The concept of “applicability” is closely related to generalizability, but places greater emphasis on nonbiological factors. Dans and colleagues²⁸ suggest six points to consider divided into categories of biologic, social and economic, and epidemiologic when considering whether a trial results applies to a specific patient. The biological factors relate to differences in the disease characteristics between the

study population and an individual patient. These factors may either enhance or diminish the expected response to treatment. The social and economic factors have to do with the likelihood of patient compliance and the ability of the provider to deliver the intervention in the same manner as in the trial. The last set of factors consider whether a patient has comorbid conditions or an increased risk of adverse outcomes attributable to factors other than the biological characteristics of the disease that will affect the potential risks and benefits of treatment.

One area where there could clearly be progress is in a broader representation of patients from different racial and ethnic groups. Findings from epidemiologic studies suggest that Parkinson’s disease is approximately 1.5–2 times more common in Caucasians than in African Americans.^{29,30} However, the proportion of participants in PD trials who are African American is typically less than 5%.^{31,32} The situation is similar in trials of multiple sclerosis.³³ In other areas the balance is somewhat better, but there is still significant underrepresentation of nonwhites. The incidence of stroke³⁴ and epilepsy³⁵ is approximately the same in whites and nonwhites; however, the proportion of nonwhite subjects participating in recent stroke trials in North America is about 30%,^{18,19} and racial participation in epilepsy trials has also been an issue.³⁶ Improving demographic representation in clinical trials will require substantial commitment of time and resources but will greatly enhance the usefulness of the trial results.

The generalizability of trial results would also be enhanced by the participation of a greater variety of practice types among recruiting centers for clinical trials. Until recently, academic specialty centers with large concentrations of particular types of patients and expertise in the clinical evaluation of these patients were the only type of practice in which clinical trials were conducted. There is now more emphasis on community-based networks of clinicians working in concert with academic centers to recruit patients for clinical trials. This approach has a prominent place in the section of the NIH Roadmap (available at <http://nihroadmap.nih.gov>) that describes plans for improving the national clinical research enterprise. Such community-based networks would enhance recruitment for clinical trials and diversify the types of practice settings in which trials are conducted. Rare disorders would continue to be studied in academic centers. However, studies of more common disorders would benefit from a such diversification. A diversification of study centers would probably have a “trickle down” effect on the study population and lead to greater diversification both in terms of race and ethnicity as well as socioeconomic status of study participants.

A second major threat to the generalizability of trial results is excessive structuring of the therapeutic regimen so that the way that a therapy is delivered in a trial would

TABLE 1. *Quasi-Randomized Alternatives to Simple Randomized Trials*

Design	Assignment of Treatment Regimen
Traditional randomized trial	Participants are assigned to one treatment regimen at random at the start of the trial.
Fixed adaptive trial	Participants are assigned to one treatment regimen at random at the start of the trial. Treatment assignments include prespecified algorithms for augmenting therapy if the subject is not responding.
Randomized adaptive trial	Participants are assigned to one treatment regimen at random at the start of the trial. All subjects are re-randomized at regular intervals in the trial. Randomization is weighted so that subjects who are responding to treatment are more likely to remain on their current treatment and subjects who are not responding are more likely to be assigned to an alternative treatment.
Randomized consent trial	Participants are assigned to one treatment regimen at random at the start of the trial. Participants are then given the option of staying in their randomized group or choosing the treatment they prefer.
Partially randomized patient preference trial	Participants are given the option of choosing the treatment they prefer or being randomized.

Each of the alternatives to the randomized trial design allows for more naturalistic treatment, and may increase participation by providing greater control over treatment. However, alternative designs limit the control of bias provided by a traditional randomized trial and require larger sample sizes.

be difficult to implement in practice. A group of novel trial designs have been proposed to address this problem and also to enhance the participation of subjects who might not ordinarily participate in a clinical trial because they are uncomfortable with the idea of being randomly assigned to therapy. These alternative designs include: 1) fixed adaptive designs, 2) randomized adaptive designs, 3) randomized consent designs, and 4) partially randomized patient preference designs.³⁷ In a fixed adaptive design trial, subjects are randomized to treatment arms that include pre-set algorithms for treatment adjustment. The use of a second "rescue" medication if a subject fails to respond to the initial therapy would be an example of a fixed adaptive design. In randomized, adaptive trials subjects are randomized initially and then at regular intervals. At each new randomization the likelihood of being randomized to a given therapy is partially determined by previous outcomes and patient and provider preferences (biased coin toss). In a randomized, consent design, subjects are randomized and then given the option of continuing on their assigned treatment or switching to the alternative based on their preference. A partially randomized, patient preference design trial is somewhat similar in that subjects may choose their therapy. In a trial of this design, subjects are given the choice between choosing a given therapy or allowing themselves to be randomized. The advantages of these designs is that they may be more attractive to patients who wish to retain some control over their treatment and that some paradigms, particularly the fixed adaptive and randomized adaptive designs, more closely reflect the medication adjustment that occurs in practice. These advantages must be balanced against the disadvantages that these alternative designs introduce limitations on the capacity

of randomization to control bias and that they generally require larger sample sizes than simple randomized trials (Table 1).

The large, simple trial paradigm is an alternative to these novel designs. The use of the large, simple trial has been pioneered in cardiovascular disease,^{38,39} and there are plans to use this paradigm in a long-term trial of potentially neuroprotective therapies for Parkinson's disease.⁴⁰ In a large, simple trial, treatment is administered in a naturalistic setting that is intended to mirror standard practice, and the outcomes measured are simple and available in the typical practice setting. Potential problems such as noncompliance with treatment, use of concomitant medications with effects similar to the experimental treatment, and response variability attributable to differences between practice styles at participating physicians are overcome by the large size of such trials. The advantages of a large, simple trial is its inclusiveness, relevance to the usual practice setting, and naturalistic approach to therapy.

PRACTICAL CLINICAL TRIALS

To choose among potential treatment regimens, physicians and other medical decision-makers need high-quality evidence based on head-to-head comparisons of clinically relevant alternatives. This type of trial has been referred to as a Practical Clinical Trial (PCT).⁴¹ The other characteristic features of practical clinical trials are, as follows: 1) they include a diverse population of study participants, 2) make use of a variety of practice settings, and 3) collect data across a broad range of health outcomes (Table 2). In spite of the inherent usefulness of PCTs, they have been relatively rare compared to placebo-

TABLE 2. *Characteristics of Practical Clinical Trials*

-
- Interventions compared in the trial are clinically relevant alternatives
 - Participants are diverse and reflect underlying affected population with the disease
 - Participants come from a heterogeneous group of practice settings and geographic locations
 - Endpoints of the trial reflect a broad range of meaningful clinical outcomes
-

bo-controlled trials. There are several reasons for this disparity. First, it is not the primary mission of major funding sources for clinical trials to conduct PCTs. The pharmaceutical industry needs placebo-controlled trials to produce the proof-of-principal evidence that is required by the relevant regulatory groups. Based on the Food and Drug Administration classification system, most practical trials would be considered phase IV trials, or postmarketing trials.⁴² The mission of the National Institutes of Health (NIH) is primarily to promote biomedical discovery, and so, like the pharmaceutical industry, the NIH is more likely to support placebo-controlled studies.⁴³ In addition, PCTs may be expensive to conduct because of large sample size requirements and extended follow-up. Several recent PCTs enrolled over 1000 subjects and cost tens of millions of dollars to complete.³⁹ High costs of PCTs compounds the problem of relative lack of interest from traditional funding sources.

Nonetheless, there have been a number of PCTs conducted in various areas of neurology. The Veterans Administration cooperative epilepsy study comparing carbamazepine to valproic acid to placebo is a classic example of a PCT.⁴⁴ More recently, several stroke trials have compared clinically relevant endpoints. The WARSS study,¹⁸ the Stroke Prevention in Atrial Fibrillation trials,^{45,46} and a trial comparing aspirin to a fixed combination of aspirin and dipyridimole⁴⁷ compared antiplatelet therapy to anticoagulation and thus provided badly needed evidence for a common clinical decision. A similar study comparing warfarin to aspirin for symptomatic intracranial atherosclerosis⁴⁸ has been recently completed and will also provide evidence to address that important clinical question. Although trials of immunomodulatory drugs for MS have mostly been placebo-controlled, there is at least one study⁴⁹ that directly compared every-other-day interferon β -1b to weekly interferon β -1a. This study showed approximately a 25% relative increase in the probability of remaining relapse-free for subjects receiving every-other-day therapy without significant differences in adverse events or patient compliance. In spite of these examples, there is a great need for more PCTs in neurology. It is easy to see the value of practical trials. PCTs are a crucial ingredient for

developing practice guidelines and quality indicators. They are also necessary for formulating evidence-based coverage policies for public and private insurers. This is especially true as large numbers of traditional, placebo-controlled trials continually expand the range of high-cost new technologies that are available to physicians and patients.

EXPANDING THE USES OF CLINICAL TRIALS

The vast majority of clinical trials test one chemotherapeutic strategy against another. However, there are several other categories that could, and probably should be subjected to the crucible of a clinical trial. It is frequently noted that surgical therapies are rarely tested in clinical trials. The ethics of surgical trials have been the source of lively debate^{50,51} and are discussed in this issue of *NeuroRx*[®]. In fact, there have been a number of surgical trials for neurological diseases, including two placebo-controlled trials of implantation of fetal dopaminergic tissue for advanced Parkinson's disease^{11,12} and randomized trials of vagus nerve stimulation⁵² and brain resection for intractable epilepsy.²¹ It is probably not practical or ethical to subject all surgical interventions to randomized trials; however, there are areas that probably should be evaluated using this format. One such area is the use of intra-arterial stents for carotid and intracranial atherosclerotic disease. Another possibility is a trial of earlier gastrostomy for patients with bulbar dysfunction attributable to motor neuron disease. Subjecting surgical interventions to the same standards as medical therapies would provide needed evidence on how to use these treatments that are both potentially very effective and also highly resource-intensive.

Another area that has been relatively neglected by clinical trials is health services interventions. This category includes randomized trials to evaluate the pharmacoeconomics of emerging therapies and trials of interventions designed to improve the delivery of healthcare in which doctors or other components of the healthcare delivery system, rather than patients, are the unit of measurement. Increasingly, pharmacoeconomic analyses are conducted along with phase III clinical trials of neurological therapies. The results of such an analysis have been published for the use of a dopamine agonist for Parkinson's disease,⁵³ cholinesterase inhibitor therapy for dementia,⁵⁴ and immune-modulatory therapy for MS,^{55,56} among others. In addition, the methods for conducting pharmacoeconomic trials alongside clinical trials are becoming more standardized.⁵⁷ However, pharmacoeconomic analyses are often underpowered and the combination of *post hoc* analysis plans and funding by sources with a vested interest in a favorable pharmacoeconomic outcome often limits the impact of these studies.

Trials of interventions directed at doctors have been quite rare in neurology. In one example, practices across New York state were randomly assigned to receive either an evidence-based educational intervention on care of patients with dementia or a standard care intervention.⁵⁸ Following this intervention, charts were reviewed for the two groups. The results showed that practices that received the intervention had greater adherence to guidelines, and avoided costly and unnecessary tests to a greater extent than practices that received the placebo intervention. The tools to evaluate interventions directed at doctors or other components of the health care system are becoming increasingly available. Quality indicators to measure the process of delivering care are under development for several neurological diseases.^{59,60} In addition, electronic medical records are becoming increasingly available, and make the chart review process necessary to measure the impact of an intervention on healthcare delivery more feasible than in the past.

Evaluation of diagnostic tests is another area where clinical trials should be put to greater use. It is considered the norm to measure the utility of a diagnostic test by assessing the characteristics of the test, including sensitivity and specificity. Standards for performing this type of evaluation have been promulgated, further entrenching the distinction between diagnostic tests and more directly therapeutic interventions.⁶¹ However, the distinction between diagnosis and therapy is somewhat artificial. Both tests and treatments are interventions intended to improve health outcomes; both have the potential for unintended consequences (side effects, incorrect results). It is logical that tests should be evaluated in the same manner as therapies. The concept of randomized controlled trials for diagnostic tests has been in the medical literature for a number of years,⁶² and some randomized trials of diagnostic tests have been performed in areas such as magnetic resonance imaging for low back pain.⁶³ The use of diagnostic testing, particularly imaging, is extremely prevalent in neurological disease. There is genuine uncertainty, if not downright skepticism in some cases, regarding the impact that additional testing has on health outcomes. Trials of testing strategies for neurological disease would help to clarify the use of emerging and existing diagnostic modalities.

DESIGNS TO EVALUATE TREATMENT EFFECTS IN CHRONIC DISEASES

One area that deserves special attention is the problem of designing clinical trials to assess treatments of chronic neurological diseases. Clearly, a substantial group of neurological disorders are chronic in nature. This group includes such conditions as Alzheimer's disease, Parkinson's disease, and multiple sclerosis. Paradoxically, the duration of trials of new therapies for these conditions

has typically been fairly brief, often 6 months to 1 year in duration. This disparity between the course of these diseases and the length of the trials is an obvious weakness in trial design. However, practical imperatives including the reluctance of many trial participants to be exposed to placebo for long periods of time, and the desire on the part of academic investigators and pharmaceutical industry sponsors to have results in a reasonable period of time necessitate short-term trials. Several strategies have been suggested to estimate long-term effects from short-term trials.

The use of surrogate endpoints, either clinical or biological, has been one predominant strategy. Although it is common to think of biomarkers as surrogate endpoints, any intermediate measure that is substituted for a true, meaningful outcome is a surrogate. Accordingly, an incremental change in the Kurtzke Expanded Disability Status Scale (EDSS)⁶⁴ or the Unified Parkinson's Disease Rating Scale (UPDRS)⁶⁵ in Parkinson's disease are really surrogate endpoints. Likewise, reduction in seizure frequency is a surrogate measure for the impact that chronic epilepsy has on quality-of-life and quality-adjusted life expectancy. Because measures like the UPDRS or the EDSS are intermediate clinical outcomes, they must satisfy two conditions to be valid: changes in the intermediate outcomes must predict important clinical outcomes in the future, and they must capture the net effects of the specific intervention in a given trial.⁶⁶

Although many commonly used clinical endpoints are actually surrogates, most of the recent debate about the use of surrogate measures focuses on biomarkers. In many cases the ability to measure these biomarkers is the result of recent technological innovation. The attractiveness of the new technology is sometimes nearly irresistible, and one may be tempted to forget that the new, technologically intensive biomarkers must meet the same standards of validity as any other surrogate to be accepted. Given that proper validation studies are performed, there is great potential in these emerging technologies.

Magnetic resonance imaging (MRI) measures of plaque burden were an important surrogate measure in early trials of immune-modulatory therapies for multiple sclerosis.⁶⁷ However, standard MRI appears somewhat quaint by contemporary standards, and has been criticized as a measure of the chronic effects of demyelinating disease, and newer MRI techniques that estimate atrophy secondary to chronic demyelination are believed to be more meaningful.⁶⁸ In the area of Parkinson's disease, fluorodopa PET,⁶⁹ and β -CIT SPECT⁷⁰ have been used as indirect measures of dopaminergic function. Fluorodopa PET measures metabolic (decarboxylase) activity and β -CIT measures dopamine transporter binding. A number of nonimaging potential biomarkers have been identified for use in clinical studies of Alzheimer's disease. Several of these markers, such as tau protein levels,

must be measured in the CSF.⁷¹ The requirement of a lumbar puncture would certainly limit the acceptance of such tests by physicians and patients. However, other potential biomarkers for Alzheimer's disease, such as isoprostanes, which are products of lipid peroxidation, can be accessed from serum and urine.⁷² Such biomarkers that are relatively inexpensive to measure and require no special equipment (like a PET scanner) to measure represent a substantial logistical advance.

A second approach to identifying durable effects over a short period of time has been the development of trial paradigms that are specifically designed to separate short-term and long-term effects. An example of such a design is the randomized, delayed-start trial.⁷³ In a randomized, delayed-start trial, one group of subjects receives treatment immediately after randomization, and the second group receives placebo initially, and active treatment is delayed. Variations on this paradigm to test dose-response relationships are possible. A trial could have many arms with staggered initiation of treatment, or the titration from a subtherapeutic dose to an effective dose could be staggered among treatment groups. The idea behind a randomized-start design is that if a treatment has a chronic (neuroprotective) effect, exposure to this treatment over a longer period of time will produce greater benefit. This design has been used in one recent trial of the monoamine oxidase type-B inhibitor, rasagiline, in patients with early PD.⁷⁴ This analysis showed a small but statistically detectable difference for the group receiving active treatment for a longer period of time. In the case of Parkinson's disease, the advantage of a randomized delayed-start design is that it controls for short-term symptomatic effects of treatments that are frequently present in compounds that may also have long-term, disease-modifying effects. The randomized, delayed-start paradigm has also been employed in trials of immune-modulatory drugs for multiple sclerosis.⁷⁵ In the case of MS, the randomized, delayed-start paradigm can be used to show that relapses are not merely delayed, but actually prevented.

Although these innovations in disease measurement and trial design do allow for the generation of evidence with implications for long-term outcomes in chronic disease, it remains the case that intermediate endpoints are surrogates for long-term outcomes and short-term trials are surrogates for long-duration trials. Ultimately, the needed evidence on the long-term effects of treatment in chronic neurological disease may need to come from trials that observe subjects over a long period of time. This approach is being implemented as part of a long-term effort by the NIH to identify neuroprotective compounds for PD in a "large simple trial" design. This time-frame is in keeping with the natural history of Parkinson's disease, and will hopefully produce results that lend themselves to straightforward interpretation. The

success of the effort will depend on the willingness of subjects to accept placebo for a prolonged period of time and adequate funding and resolve on the part of investigators.

The randomized, controlled clinical trial is likely to remain the gold standard for evidence for clinical decision-making in neurology. Although observational methods and systematic reviews are clearly useful, neither provides the control of confounding and bias that insures the internal validity of a randomized trial. Specifically because of the central importance of clinical trials, it is crucial to continue to work to reduce their inherent limitations, including uncertain generalizability, and to expand the uses of the randomized clinical trial paradigm to areas beyond proving biological activity. Making progress in these areas will be the challenge for clinical trials in neurology as we move into the future.

Acknowledgments: This work was supported by Grant K-08 HS00004 from the Agency for Healthcare Research and Quality.

I thank Drs. Scott Kasner, Jacqueline French, Steven Galetta, and Laura Balcer for helpful suggestions.

REFERENCES

1. Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ* 327:1459–1461, 2003.
2. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 270:2598–2601, 1993.
3. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. The North American Symptomatic Carotid Endarterectomy Trial Collaborators. *N Engl J Med* 325:445–453, 1991.
4. Stroke Prevention in Atrial Fibrillation Investigators. Adjusted-dose warfarin versus low-intensity, fixed-dose warfarin plus aspirin for high-risk patients with atrial fibrillation: Stroke Prevention in Atrial Fibrillation III randomised clinical trial. *Lancet* 348:633–638, 1996.
5. Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med* 342:1907–1909, 2000.
6. Anderson GL, Judd HL, Kaunitz AM, Barad DH, Beresford SA, Pettinger M et al. Effects of estrogen plus progestin on gynecologic cancers and associated diagnostic procedures: the Women's Health Initiative randomized trial. *JAMA* 290:1739–1748, 2003.
7. Grady D, Herrington D, Bittner V, Blumenthal R, Davidson M, Hlatky M et al. Cardiovascular disease outcomes during 6.8 years of hormone therapy: Heart and Estrogen/progestin Replacement Study follow-up (HERS II). *JAMA* 288:49–57, 2002.
8. Austin PC, Mamdani MM, Tu K, Jaakkimainen L. Prescriptions for estrogen replacement therapy in Ontario before and after publication of the Women's Health Initiative Study. *JAMA* 289:3241–3242, 2003.
9. Freeman TB, Olanow CW, Hauser RA, Nauert GM, Smith DA, Borlongan CV et al. Bilateral fetal nigral transplantation into the postcommisural putamen in Parkinson's disease. *Ann Neurol* 38: 379–388, 1995.
10. Lindvall O, Sawle GV, Widner H. Evidence for long-term survival and function of dopaminergic grafts in progressive Parkinson's disease. *Ann Neurol* 35:172–180, 1994.
11. Freed CR, Greene PE, Breeze RE, Tsai W-Y, DuMouchel W, Kao R et al. Transplantation of embryonic dopamine neurons for severe Parkinson's disease. *N Engl J Med* 344:710–719, 2000.

12. Olanow CW, Goetz CG, Kordower JH, Stoessl AJ, Sossi V, Brin MF et al. A double-blind controlled trial of bilateral fetal nigral transplantation in Parkinson's disease. *Ann Neurol* 54:403-414, 2003.
13. Furlan AJ, Cavalier SJ, Hobbs RE. Hemorrhage and anticoagulation after nonseptic embolic brain infarction. *Neurology* 32:280-282, 1982.
14. Adams HP Jr, Brott TG, Crowell RM, Furlan AJ, Gomez CR, Grotta J et al. Guidelines for the management of patients with acute ischemic stroke. A statement for healthcare professionals from a special writing group of the Stroke Council, American Heart Association. *Circulation* 90:1588-1601, 1994.
15. The International Stroke Trial (IST): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19435 patients with acute ischaemic stroke. International Stroke Trial Collaborative Group. *Lancet* 349:1569-1581, 1997.
16. Johnston SC. Identifying confounding by indication through blinded prospective review. *Am J Epidemiol* 154:276-284, 2001.
17. Guyatt GH, Cook DJ, Sackett DL, Eckman M, Pauker SG. Grades of recommendation for antithrombotic agents. *Chest* 114:441S-444S, 1998.
18. Mohr JP, Thompson JL, Lazar RM, Levin B, Sacco RL, Furie KL et al. A comparison of warfarin and aspirin for the prevention of recurrent ischemic stroke. *N Engl J Med* 345:1444-1451, 2001.
19. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and the Stroke rt-PA Stroke Study Group. *N Engl J Med* 333:1581-1587, 1995.
20. CAST: randomised placebo-controlled trial of early aspirin use in 20,000 patients with acute ischaemic stroke. CAST (Chinese Acute Stroke Trial) Collaborative Group. *Lancet* 349:1641-1649, 1997.
21. Wiebe S, Blume WT, Girvin JP, Eliasziw M. Effectiveness and Efficiency of Surgery for Temporal Lobe Epilepsy Study Group. A randomized, controlled trial of surgery for temporal-lobe epilepsy. *N Engl J Med* 345:311-318, 2001.
22. Interferon beta-1b in the treatment of MS: final outcome of the randomized controlled trial. The IFNB Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group. *Neurology* 45:1277-1285, 1995.
23. Jacobs LD, Beck RW, Simon JH, Kinkel RP, Brownscheidle CM, Murray TJ et al. Intramuscular interferon beta-1a therapy initiated during a first demyelinating event in multiple sclerosis. CHAMPS Study Group. *N Engl J Med* 343:898-904, 2000.
24. Johnson KP, Brooks BR, Cohen JA, Ford CC, Goldstein J, Lisak RP et al; Copolymer 1 Multiple Sclerosis Study Group. Extended use of glatiramer acetate (Copaxone) is well tolerated and maintains its clinical effect on multiple sclerosis relapse rate and degree of disability. *Neurology* 50:701-708, 1998.
25. Fazekas F, Barkhof F, Filippi M, Grossman RI, Li DK, McDonald WI et al. The contribution of magnetic resonance imaging to the diagnosis of multiple sclerosis [review]. *Neurology* 53:448-456, 1999.
26. Beck RW, Cleary PA, Anderson MM Jr, Keltner JL, Shults WT, Kaufman DI et al. A randomized, controlled trial of corticosteroids in the treatment of acute optic neuritis. The Optic Neuritis Study Group. *N Engl J Med* 326:581-588, 1992.
27. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. *JAMA* 282:771-778, 1999.
28. Dans AL, Dans LF, Guyatt GH, Richardson S. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. Evidence-Based Medicine Working Group. *JAMA* 279:545-549, 1998.
29. Mayeux R, Marder K, Cote LJ, Denaro J, Hemenegildo N, Mejia H et al. The frequency of idiopathic Parkinson's disease by age, ethnic group, and sex in northern Manhattan, 1988-1993. *Am J Epidemiol* 142:820-827, 1995.
30. Van Den Eeden SK, Tanner CM, Bernstein AL, Fross RD, Leim-peter A, Bloch DA et al. Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. *Am J Epidemiol* 157:1015-1022, 2003.
31. Parkinson Study Group. Pramipexole vs Levodopa as initial treatment for Parkinson's disease. *JAMA* 284:1931-1938, 2000.
32. Adler CH, Singer C, O'Brien C, Hauser RA, Lew MF, Marek KL et al. Randomized, placebo-controlled study of tolcapone in patients with fluctuating Parkinson disease treated with levodopa-carbidopa. Tolcapone Fluctuator Study Group III. *Arch Neurol* 55:1089-1095, 1998.
33. Hoggancamp WE, Rodriguez M, Weinshenker BG. The epidemiology of multiple sclerosis. *Mayo Clinic Proc* 72:871-878, 1997.
34. Warlow C, Sudlow C, Dennis M, Wardlaw J, Sandercock P. Stroke. *Lancet* 362:1211-1224, 2003.
35. Miller LL, Pellock JM, Boggs JG, DeLorenzo RJ, Meyer JM, Corey LA. Epilepsy and seizure occurrence in a population-based sample of Virginian twins and their families. *Epilepsy Res* 34:135-143, 1999.
36. Berg AT, Vickrey BG, Langfitt JT, Sperling MR, Walczak TS, Shinnar S et al. The multicenter study of epilepsy surgery: recruitment and selection for surgery. *Epilepsia* 44:1425-1433, 2003.
37. TenHave TR, Coyne J, Salzer M, Katz I. Research to improve the quality of care for depression: alternatives to the simple randomized clinical trial. *Gen Hosp Psych* 25:115-123, 2003.
38. Topol EJ, Califf RM, Van de Werf F, Simoons M, Hampton J, Lee KL et al. Perspectives on large-scale cardiovascular clinical trials for the new millennium. The Virtual Coordinating Center for Global Collaborative Cardiovascular Research (VIGOUR) Group. *Circulation* 95:1072-1082, 1997.
39. Wright JT Jr, Cushman WC, Davis BR, Barzilay J, Colon P, Egan D et al. The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT): clinical center recruitment experience. *Control Clin Trials* 22:659-673, 2001.
40. Heemskerk J, Tobin AJ, Ravina B. From chemical to drug: neurodegeneration drug screening and the ethics of clinical trials. *Nat Neurosci [Suppl 5]:*1027-1029, 2002.
41. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 290:1624-1632, 2003.
42. Investigational New Drug Application. Code of Federal Regulations Title 21 Part 5, pp 300-499. Revised as of 4-1-1999. Washington, DC: Government Printing Office, 1999.
43. Sung NS, Crowley WF Jr, Genel M, Salber P, Sandy L, Sherwood LM et al. Central challenges facing the national clinical research enterprise. *JAMA* 289:1278-1287, 2003.
44. Mattson RH, Cramer JA, Collins JF. A comparison of valproate with carbamazepine for the treatment of complex partial seizures and secondarily generalized tonic-clonic seizures in adults. The Department of Veterans Affairs Epilepsy Cooperative Study 264 Group. *N Engl J Med* 327:765-771, 1992.
45. Preliminary report of the stroke prevention in atrial fibrillation study. *N Engl J Med* 322:863-868, 1990.
46. Warfarin versus aspirin for prevention of thromboembolism in atrial fibrillation. Stroke Prevention in Atrial Fibrillation II Study. *Lancet* 343:687-691, 1994.
47. Diener HC, Cunha L, Forbes C, Sivenius J, Smets P, Lowenthal A. European Stroke Prevention Study. 2. Dipyridamole and acetylsalicylic acid in the secondary prevention of stroke. *J Neurol Sci* 143:1-13, 1996.
48. Benesch CG, Chimowitz MI. Best treatment for intracranial arterial stenosis? 50 years of uncertainty. The WASID Investigators. *Neurology* 55:465-466, 2000.
49. Durelli L, Verdun E, Barbero P, Bergui M, Versino E, Ghezzi A et al. Every-other-day interferon beta-1b versus once-weekly interferon beta-1a for multiple sclerosis: results of a 2-year prospective randomised multicentre study (INCOMIN). *Lancet* 359:1453-1460, 2002.
50. Macklin R. The ethical problems with sham surgery in clinical research. *N Engl J Med* 341:992-996, 1999.
51. Freeman TB, Vawter DE, Godbold JH, Hauser RA, Goetz CG et al. Use of placebo surgery in controlled trials of a cellular-based therapy for Parkinson's disease. *N Engl J Med* 341:988-992, 1999.
52. Handforth A, DeGiorgio CM, Schachter SC, Uthman BM, Naritoku DK, Tecoma ES et al. Vagus nerve stimulation therapy for partial-onset seizures: a randomized active-control trial. *Neurology* 51:48-55, 1998.

53. Hoerger TJ, Bala MV, Rowland C, Greer M, Chrischilles EA, Holloway RG. Cost effectiveness of pramipexole in Parkinson's disease in the US. *Pharmacoeconomics* 14:541–557, 1998.
54. Clegg A, Bryant J, Nicholson T, McIntyre L, De Broe S, Gerard K et al. Clinical and cost-effectiveness of donepezil, rivastigmine, and galantamine for Alzheimer's disease. A systematic review. *Int J Technol Assess Health Care* 18:497–507, 2002.
55. Touchette DR, Durgin TL, Wanke LA, Goodkin DE. A cost-utility analysis of mitoxantrone hydrochloride and interferon beta-1b in the treatment of patients with secondary progressive or progressive relapsing multiple sclerosis. *Clin Ther* 25:611–634, 2003.
56. Kobelt G, Jonsson L, Miltenburger C, Jonsson B. Cost-utility analysis of interferon beta-1B in secondary progressive multiple sclerosis using natural history disease data. *Int J Technol Assess Health Care* 18:127–138, 2002.
57. Glick H. Strategies for economic assessment during the development of new drugs. *Drug Inf J* 29:1391–1403, 1995.
58. Gifford DR, Holloway RG, Frankel MR, Albright CL, Meyerson R, Griggs RC et al. Improving adherence to dementia guidelines through education and opinion leaders. A randomized, controlled trial. *Ann Intern Med* 131:237–246, 1999.
59. Cheng EM, Siderowf A, Swartztrauber K, Eisa M, Lee M, Vickrey BG. Development of quality of care indicators for Parkinson's disease. *Mov Disord* 19:136–150, 2004.
60. Holloway RG, Vickrey BG, Benesch C, Hinchey JA, Bieber J; National Expert Stroke Panel. Development of performance measures for acute ischemic stroke. *Stroke* 32:2058–2074, 2001.
61. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 138:40–44, 2003.
62. Fineberg HV, Bauman R, Sosman M. Computerized cranial tomography. Effect on diagnostic and therapeutic plans. *JAMA* 238:224–227, 1977.
63. Jarvik JG, Hollingworth W, Martin B, Emerson SS, Gray DT, Overman S et al. Rapid magnetic resonance imaging vs radiographs for patients with low back pain: a randomized controlled trial. *JAMA* 289:2810–2818, 2003.
64. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 33:1444–1452, 1983.
65. Fahn S, Elton RL. Unified Parkinson's disease rating scale. In: Recent developments in Parkinson's disease (Fahn S, Marsden CD, Calne D, Goldstein M, eds), pp 153–164. Florham Park, NJ: Macmillan Health Care Information, 1987.
66. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? [review] *Ann Intern Med* 125:605–613, 1996.
67. Stone LA, Albert PS, Smith ME, DeCarli C, Armstrong MR, McFarlin DE et al. Changes in the amount of diseased white matter over time in patients with relapsing-remitting multiple sclerosis. *Neurology* 45:1808–1814, 1995.
68. Bergers E, Bot JC, De Groot CJ, Polman CH, Nijeholt GJ, Castelijns JA et al. Axonal damage in the spinal cord of MS patients occurs largely independent of T2 MRI lesions. *Neurology* 59:1766–1771, 2002.
69. Morrish PK, Sawle GV, Brooks DJ. Regional changes in [18F]dopa metabolism in the striatum in Parkinson's disease. *Brain* 119:2097–2103, 1996.
70. Seibyl JP, Marek K, Sheff K, Zoghbi S, Baldwin RM, Charney DS et al. Iodine-123-beta-CIT and iodine-123-FPCIT SPECT measurement of dopamine transporters in healthy subjects and Parkinson's patients. *J Nucl Med* 39:1500–1508, 1998.
71. Clark CM, Xie S, Chittams J, Ewbank D, Peskind E, Galasko D et al. Cerebrospinal fluid tau and beta-amyloid: how well do these biomarkers reflect autopsy-confirmed dementia diagnoses? *Arch Neurol* 60:1696–1702, 2003.
72. Pratico D, Clark CM, Lee VM, Trojanowski JQ, Rokach J, FitzGerald GA. Increased 8,12-iso-iPF2alpha-VI in Alzheimer's disease: correlation of a noninvasive index of lipid peroxidation with disease severity. *Ann Neurol* 48:809–812, 2000.
73. Leber P. Slowing the progression of Alzheimer disease: methodologic issues. *Alzheimer Dis Assoc Disord* 11:S10–S21, 1997.
74. Parkinson Study Group. A controlled, randomized, delayed-start study of rasagiline in early Parkinson's disease. *Arch Neurol* 61:561–566, 2004.
75. PRISMS Study Group and the University of British Columbia MS/MRI Analysis Group. PRISMS-4: long-term efficacy of interferon-beta-1a in relapsing MS. *Neurology* 56:1628–1636, 2001.