

Use of this content is subject to the [Terms and Conditions](#) of the MD Consult web site.

Neurologic Clinics

Volume 14 • Number 2 • May 1996

Copyright © 1996 W. B. Saunders Company

451

NEUROEPIDEMIOLOGY

CLINICAL RESEARCH IN NEUROLOGY

From Observation to Experimentation

Frederick J. Marshall MD

Karl Kieburtz MD, MPH

Michael McDermott PhD

Roger Kurlan MD

Ira Shoulson MD

Department of Neurology, University of Rochester School of Medicine and Dentistry, Rochester, New York

This work was supported by Training Grant No. NS-07338 from the National Institute of Neurological Disorders and Stroke.

Address reprint requests to

Frederick J. Marshall, MD

Department of Neurology

Strong Memorial Hospital

601 Elmwood Avenue, Box 673

Rochester, NY 14642

The historical development of neurology has hinged largely on keen observation and the ability of skilled clinicians to draw inferences from experience. Incremental additions to our understanding of the nervous system traditionally have emanated from two sources: the bench and the bedside. Just as technologic sophistication in basic research has evolved over the past 200 years, so too have there been major advances in clinical research methodology. If a stable bridge is to be built between new insights gained in the laboratory and new interventions applied at the bedside, this bridge must be designed according to sound research principles. These principles form the basis of the emerging science of experimental therapeutics. Application of these principles in the context of clinical research in neurology forms the basis of this review.

HISTORICAL BACKGROUND

Only within the past half century has the randomized controlled trial become the gold standard in the evaluation of therapeutic interventions. Prior to World War II, it was common for therapies to enter into widespread use without being subjected to the rigors of the scientific method. Breakthroughs in clinical medicine naturally unfolded along more haphazard lines, and systematic

452

biases crept into our understanding of diseases and therapies. [9] After 1938, federal law in the United States required

pharmaceutical companies to document toxicity experiments in animals, but it was not until the teratogenic catastrophe of thalidomide in the 1960s that legislators and regulators in this country and in Britain began to insist on well-controlled trials prior to drug approval. [52]

Advances in many areas of the basic and applied sciences throughout the first half of the twentieth century set the stage for the development of new clinical therapies. The evolution of therapies in Parkinson's disease (PD) is illustrative. Charcot [15] and his students long had recognized the utility of the belladonna alkaloids in the treatment of parkinsonian tremor. In the 1940s, organic chemists were able to synthesize novel analogues of these alkaloids, providing anticholinergic treatments that were tolerated better. [18] These agents remained the therapeutic mainstay in PD until the advent of levodopa in the 1960s.

Fluorescence microscopy and new biochemical assays made it possible for researchers to demonstrate the distribution of dopamine-containing neurons in the CNS by the end of the 1950s. [39] The suppression of dopamine levels in reserpine-induced parkinsonism and its reversal with administration of the precursor molecule 3,4-dihydroxyphenylalanine (dopa) formed the basis of the dopamine hypothesis of PD. [12] The original trials of levodopa in patients, however, were discouraging. Peripheral decarboxylation of dopa to dopamine resulted in excessive nausea. More than a dozen inconclusive studies--most small, unblinded, and without benefit of placebo control--were performed before 1967. [34]

Fehling [22] reported the first double-blind study of levodopa in 1966. She examined intravenous administration of the drug in 27 patients, concluding that the only difference between levodopa and saline was that the former provoked nausea in a third of patients. [22] The fact that levodopa went on to revolutionize the treatment of PD is due to the persistence of Cotzias and coworkers, who eventually managed to give racemic dopa in concentrations that were an order of magnitude greater than most prior studies had used. [17] This was accomplished by slowly increasing the dosage and allowing patients to become tolerant to the drug's emetic side effect. These researchers were able to achieve remarkable improvement in patients' symptoms and signs. Subsequent large, carefully controlled studies confirmed these results and established the efficacy of levodopa in PD. [36] [81]

The development of levodopa therapy in PD nicely demonstrates three important principles in the conduct and interpretation of clinical trials: (1) basic research can and should serve as a compelling starting point in the generation of clinical hypotheses; (2) regardless of how compelling the theoretic implications of basic research, experimental therapies ultimately must stand or fall on their own merit; the clinical investigator always should be willing to abandon the high road of theory for the low road of practice by answering a deceptively simple question: "Is this agent safe and effective at this dosage in these patients?"; and (3) clinical trials represent potent means for clarifying clinical outcomes, but they are generally inadequate for addressing questions of pathophysiology or the mechanism whereby a given intervention gives rise to a specific result. The possibility that a drug is acting in some unforeseen fashion or that a different (perhaps untested) dosage regimen might exert qualitatively as well as quantitatively different effects must be borne in mind always. Outcomes of clinical trials should not be used to settle underlying questions of pathophysiology. Just as basic science generates hypotheses for clinical application, clinical trials generate hypotheses for basic investigation. Clinical trials do

453

not prove or disprove hypotheses about basic mechanisms any more than basic research proves or disproves the pragmatic conclusions of applied clinical science.

OBSERVATIONAL STUDIES

Underlying Principles

Clinical studies may be divided broadly into two types: observational and interventional. [20] In observational studies, the investigator plays a relatively passive role, gathering impressions, carefully documenting clinical outcomes, and applying appropriate statistical methods to draw conclusions from the data. Interventional studies, on the other hand, involve an active intervention. Typically, the intervention is applied to individuals in one group and withheld from those in a parallel group. Formalized design methods have been developed to help ensure that these interventional clinical experiments do not suffer from bias (i.e., that there are no systematic differences in the way groups of subjects are selected and treated or in the way outcomes are measured and interpreted). The benefit of the experimental study lies in its ability to facilitate causal

inference about the effects of an intervention by ensuring that bias is minimized--that groups being compared indeed are comparable in every respect except for the intervention.

Nonetheless, observational studies remain useful on several grounds. Such studies may provide information about cause or pathogenesis. They are particularly useful when ethical constraints preclude intervening in ways that alter the historical antecedents of a disease. For example, one could not determine ethically whether head trauma causes epilepsy by conducting an active interventional experiment. Further, the principle of clinical *equipoise*--the requirement that investigators must be truly uncertain about the relative merits of their experimental intervention--places ethical constraints on experimental trials and may prompt investigators to pursue an observational study. [23] Finally, logistic considerations may preclude an experimental intervention for reasons of cost, time, or availability of subjects.

The most significant problem with observational studies, whether they are prospective or retrospective, is that they are subject to bias; they cannot ensure that comparison groups indeed are comparable in all respects except for the factor of interest. This is particularly a problem when one cannot measure (or even know) all of the potentially confounding variables.

There are several standard forms of observational study in clinical medicine, most of which are well known to neurologists from their reading of the clinical literature.

Types of Observational Study

Case Reports and Case Series

The most straightforward type of observational study is a single case report or case series. From a historical perspective, case reporting long has held an important role in the development of clinical neurology. Eponymic diseases and syndromes abound. James Parkinson in 1817 reported on a series of six patients with a "shaking palsy," one of whom he had observed only from afar while

454

walking in the streets of London. [42] George Huntington in 1872 described the fundamental features of "hereditary chorea," culling from his own clinical experience and that of his father and grandfather, both of whom had preceded him in the rural practice of medicine on Long Island. [30] Case reports serve to heighten awareness of new syndromes and potential therapies, as well as generate questions for further formalized study. Although these examples are from the nineteenth century, it is clear that the case report is here to stay. Indeed, the new neurologic journal, *Neurocase*, is devoted exclusively to the publication of case reports.

Case-control (Retrospective) Studies

Unlike simple case reports, case-control studies attempt to gather evidence in support of cause-and-effect relationships. This design is *retrospective*: subjects who have a characteristic of interest (e.g., a particular disease) are identified, and their history is compared with that of controls. The control subjects tend to be selected or matched so as to be as similar as possible to the cases in respects other than the characteristic of interest. Case-control studies assess the relative frequency of specific antecedent conditions in each of the groups. Potential risk factors for the development of a disease may be investigated in this fashion. Studies in PD, for example, have looked at the presence of a positive family history of PD, exposure to herbicides, consumption of well water, smoking history, and head trauma, among other things, in patients compared with matched controls. [10] [56] [60]

In a recent article on statistical methodology in epidemiologic studies of multiple sclerosis, Hibberd [28] outlines a number of advantages and disadvantages to the case-control design. She notes the relatively low cost of performing these studies, the speed with which they can be performed, their utility in assessing rare diseases with long latencies, and their requirement for fewer subjects than prospective studies. Hibberd cautions, however, about the possibility of systematic bias involving the maintenance and scrutiny of medical records, subjects' recall of antecedent events, and investigators' selection of cases and controls. There are major difficulties relating to how exposures are defined, how subjects are matched, and how outcomes are assessed. [16] As in other observational studies, investigators in case-control paradigms are limited by their inability to match for potentially important unknown or unmeasured variables.

Another possible drawback to the retrospective design is the relative ease with which various historical antecedents may be assessed, leading to a plethora of comparisons. As Armitage and Berry ^[2] have stated,

A sufficiently assiduous search will often reveal some remarkable contrasts which have arisen purely by chance. This may not matter if scrutiny is restricted to those comparisons which the study was designed to throw light on. If, on the other hand, the data are subjected to what is sometimes called a *dredging* procedure--a search for significant contrasts which would not have been thought of initially--there is a real danger that a number of comparisons will be reported as significant, but that they will almost all have arisen by chance.

Cohort Studies

Like the case report and the case-control study, cohort studies are observational rather than interventional. Cohort studies, however, are *prospective* in that groups are defined and identified at the outset and followed over time for the development of some outcome of interest. This type of study is somewhat more

455

powerful than other observational study types because its prospective nature enables conclusions to be drawn about the relative incidence of events as they unfold. As in the case-control study, parallel groups are defined ideally to differ only in regard to some specified factor of interest. This process is intended to make the groups comparable and to eliminate biases that might undermine the validity of any causal inference.

Factors that are controlled for commonly include age, gender distribution, and overall health status. Race and socioeconomic factors also may influence outcome and often are taken into account, as are health habits, such as smoking and consumption of alcohol. In a recent study investigating the outcome of patients with motor impairment and radiographic evidence of leukoaraiosis, Inzitari and colleagues ^[31] followed patients for several years and documented the development of stroke, myocardial infarction, and death. Thirty-one cases referred for evaluation of progressive bilateral motor deficits were selected from their hospitalized patient population aged 45 years or older on the basis of characteristic periventricular white matter changes evidenced by CT scanning. Control subjects were 68 neurology patients, aged 45 years or older, admitted to the hospital during the same period whose CT scans did not show these white matter changes. The authors attempted to make the groups similar with regard to important predictors of cerebrovascular outcome, such as prior stroke and history of hypertension. They reported a 6-year cumulative stroke risk of 49% for the cases, compared with 16% for the controls. Their preliminary conclusion was that patients with motor impairment and extensive leukoaraiosis had a very poor cerebrovascular outcome.

Benefits of this type of prospective analysis include the fact that follow-up generally is more careful and complete. Rigorous criteria for evaluating risk factors and outcome measurements facilitate interpretation of results. These studies, however, require time and expense in following patients until the development of a specified outcome. Cohort studies usually are not suitable for studying rare disorders because large numbers of subjects would have to enter the cohort to supply an ample number of outcomes for analysis. As with the case-control design, bias may be extremely difficult to avoid. In the study described in the previous paragraph, the authors noted that a previous history of lacunar infarction at baseline was much more common in their cases compared with their controls. When they adjusted statistically for this covariate, the excess risk of stroke originally inferred was no longer significant.

INTERVENTIONAL STUDIES

Underlying Principles

By minimizing potential sources of systematic bias, interventional studies are better suited to assess cause-and-effect relationships. A number of key design features in experimental trials may reduce bias and thereby provide valid and useful truths about potential therapies in patients with particular diseases. These include the concepts of intervention, randomization, blinding, and control.

Intervention

The cardinal feature of clinical trials is the controlled intervention. The logic of experiment requires that one intervention be compared with an alternative (often a placebo) in order to arrive at valid conclusions. This intervention affects

456

the status quo in a manner that is under the control of the researcher, tightening the noose of logical inference about causation.

Intervention in clinical trials most often involves administration of a drug but also can involve medical procedures or devices. Examples include the DATATOP study of the use of tocopherol and selegiline HCl (Deprenyl) in early-onset PD and the NASCET study of endarterectomy in patients with symptomatic carotid stenosis. [40] [46] [47] These trials prospectively examined the outcome of patients exposed to specific interventions.

Randomization

As Pocock has pointed out, "The idea that patients should be randomly assigned to one or other form of treatment is not intuitively appealing either to the medical profession or the layman. Superficially, randomized comparison of treatments appears contrary to the need for the clinician to give every patient the best possible care and hence appears to imply a loss of freedom for both patient and clinician." [52] This concern, however, implies that the best (or worst) alternative is known already, which should not be the case if the demand for clinical equipoise has been met. The investigator must not feel that one treatment is clearly superior to another. Chalmers and others have stressed the importance of randomization as a guard against systematic bias and as a foundation for standard methods of statistical analysis. [14] [53]

By ensuring that the chances of allocation to one group or another are known and that the allocation is determined by a random mechanism, experimenters may be assured that, in the long run, potentially confounding variables are similar among the comparison groups. On the other hand, systematic nonrandom (haphazard) assignment, such as treatment allocation on the basis of which day a subject happens to be seen in clinic, invites selection bias; the investigator consciously or unconsciously may manipulate the clinic schedule. Random allocation avoids selection bias and establishes balance among groups regarding potentially confounding variables whether or not they are measured or even known.

It is important to stress, however, that this allocation represents a random process; it remains possible for groups to differ in important respects owing to chance alone, particularly if the trial does not involve large numbers of participants. Statistical methods such as analysis-of-covariance have been developed to take into account such random deviations in the comparison of treatment groups. [1]

The techniques of stratification and blocking are two important means by which investigators may avoid potentially important differences among treatment groups that may occur owing to chance allocation alone. Subjects frequently are stratified by treatment center (in multicenter trials), age, gender, or other variables known to influence outcomes strongly. Randomization occurs within the confines of these strata, thereby ensuring that patients of one type are not under- or overrepresented in a given treatment arm. Stratification can be self-defeating if the number of strata is too large; variables to be stratified, therefore, must be chosen carefully and kept to a reasonable limit.

Blocking may be thought of as stratification with regard to the time of patient entry into the trial. In a multicenter trial in which a given site is to enroll 16 patients, for example, it may be desirable to ensure that there is a balanced allocation to the various treatment arms by the time the first eight subjects are randomized. This strategy guards against the possibility of drastically unequal treatment allocations in the event of a premature termination of the study or

457

poorer than expected recruitment and makes interim analyses more feasible. The block size should be a multiple of the number of treatment arms in the study. It is important that specific information on the block size be kept from investigators so as not to compromise blindness. For the same reason, block size should not be so small as to enable investigators to deduce the treatment allocation of a given subject. Although rare, occasionally it is necessary for an investigator to break the blind on a given patient in case of a medical emergency. Doing so should not compromise the extent to which the

investigator remains blinded to the treatment status of the other patients in the study.

Blinding

Like randomization, blinding is used to prevent the introduction of systematic biases in the conduct and interpretation of clinical trials. There are several levels of blinding, ranging from unblinded, open-label trials to double-blind trials. Several authors have stressed the importance of blinding in ensuring that investigators' enthusiasm for a particular intervention does not color the assessment of that intervention. [\[21\]](#) [\[50\]](#)

In unblinded trials, both the patient and the investigator are aware of the treatment allocation. Such unblinded studies lack a strong degree of experimental validity. Both patients and investigators may overestimate treatment effects or adverse effects so that it becomes impossible to draw unbiased conclusions about efficacy and tolerability. Subjects may be enrolled selectively in a nonrandomized study on the basis of a favorable baseline prognosis. The level of ancillary care that is provided and the degree of subjectivity involved in assessing outcomes all may undermine the conclusions of an open-label trial. Even the conclusions of unblinded trials with objective outcome criteria, such as mortality, may be undermined by differential allotment to treatment arms or by bias in the selection of subjects to participate in a nonrandomized trial. In a sense, open-label trials that are not randomized or controlled (despite the fact that they entail an intervention) are observational studies; they are analogous to cohort studies in which the attribute of interest happens to be exposure to a drug or procedure rather than the presence of a risk factor.

In a recent article on the Canadian cooperative trial of cyclophosphamide and plasma exchange in multiple sclerosis, Noseworthy et al [\[41\]](#) elegantly demonstrated the critical role of blinding in arriving at valid conclusions from clinical trial data. In this trial, both blinded and unblinded neurologists assessed patients' treatment outcomes. The unblinded neurologists' scores demonstrated an apparent treatment benefit at 6, 12, and 24 months, whereas the blinded neurologists' scores did not. The authors conclude that "the bias introduced by physician unblinding would have contributed to an erroneous conclusion about treatment efficacy in what otherwise seems to have been a negative clinical trial. . . ." [\[41\]](#)

It may be extremely difficult to ensure that patients and clinicians remain blinded to treatment allocation. In the study noted in the previous paragraph, patients in the control group actually underwent sham plasma exchange. Of course, studies of surgical interventions may be more difficult to blind. In pharmacotherapeutic studies, there may be obvious side effects of the active medication, such as discoloration of the urine or flushing of the skin. The pill itself may be bitter or otherwise recognizable.

The Standards of Reporting Trials Group has called on study authors to state explicitly the methods used to maintain the blind, so that readers may get a more realistic understanding of the strength of the blinding process. [\[59\]](#) Many

458

trials have incorporated an assessment of the adequacy of blinding within the trial itself. Patients and investigators are asked to guess what treatment was used; this guess is then compared to reality after the blind is broken at the conclusion of the trial. [\[43\]](#)

Control

Only by comparing one treatment with another, either a placebo or an alternative treatment, can investigators get a clear estimate of the impact of an intervention. Many studies have relied on historical controls to provide this comparison. The use of historical controls may reflect a lack of clinical equipoise because all subjects in such trials gain access to the new treatment, which is assumed to be superior. Such studies also may be helpful when subject numbers or financial resources are limited. [\[39\]](#) Because historically controlled studies tend to overestimate treatment efficacy, negative results may be useful for screening out likely ineffective therapies from future study. Many statisticians, however, have pointed out the potential flaws of using natural history data about a disease process to establish a comparison. [\[53\]](#) The selection criteria for study participation, the quality of recorded data, the quality of ancillary support services, and the evaluation of response all are likely to be different among patients involved in the study and those serving as historical controls.

Nonetheless, practical considerations have prompted some authors to question the need for concurrent randomized controls in all clinical trials. Pradas et al argued that the natural history controlled trial in phase II studies of amyotrophic lateral sclerosis "provides a scientifically sound and cost-effective methodology," [\[54\]](#) but there are relatively few comparative data

on this point. Pocock [51] outlined several standards for the use of natural history controls, including the need for precisely defined inclusion criteria and methods of treatment and evaluation. He emphasized that natural history control data should be gathered ideally by the same investigators involved in the treatment trial.

The ethics of placebo-controlled trials in potentially fatal diseases have been explored at length. [27] [48] Lebacqz [32] underscores the importance of clinical equipoise if concurrent controls are to be used ethically in clinical trials. [32] Many areas of clinical research in neurology already have standard therapeutic options. In epilepsy, for example, most current trials are performed using add-on treatments, as withholding standard drugs clearly can lead to patient deterioration. [6] [13] Many new therapies for PD are investigated in the context of levodopa treatment. [44]

The danger of inadequate control leading to erroneous conclusions must be kept in mind always. Indeed, it may be argued that ethical considerations compel us to use proper concurrent controls in our trials to avoid the negative consequences of a false conclusion. In light of the so-called *placebo effect*--that mere participation in a trial may alter the course of an illness--no degree of detailed knowledge about the natural history of a disease process can substitute for concurrent control.

Rather than obviate the need for concurrent controls in clinical trials, detailed natural history studies can provide researchers information for adequate estimates of sample size and trial duration. [8] In trials of therapeutic agents for Huntington's disease, for example, it is critical to have an idea about how fast the disease process normally unfolds. The Huntington Study Group, a multicenter consortium of clinical and basic scientists with an interest in Huntington's disease, currently is building a large natural history database containing information on more than 1000 patients. Knowledge of the rate and variability of

459

progression of Huntington's disease over a certain period permits estimation of trial duration and of the sample size required to detect a specified treatment effect.

For example, the primary outcome measure in a planned, two-armed clinical trial is the change in subjects' total functional capacity (TFC) score [58] from baseline to 18 months. Information was obtained on 76 patients from the natural history database who had initial TFC scores greater than 6 (reflecting inclusion/exclusion criteria for the trial) and two measurements approximately 18 months apart by the same rater. Over this period, the mean \pm standard deviation of the decline in TFC was approximately 1.0 ± 1.5 units per year. This information may be used to construct [Figure 1](#), which gives the sample sizes required to detect a 57%, 50%, or 46% slowing of the progression of the disease with various degrees of power, using a two-sided *t*-test at the 5% level of significance. To illustrate, a total sample size of 296 subjects (148 per group) would provide approximately 81.5% power to detect a difference of 0.5 units per year in the mean TFC decline between the two groups. If, in truth, the treatment yields 0.5 units per year (50%) slowing of the disease process, we would have an 81.5% chance of demonstrating a statistically significant difference in the rate of disease progression using this sample size. These calculations may be refined to take into account the effect of patients who do not complete the study.

The 18-month duration of this trial was chosen because it was sufficiently long to enable detection of the effect of the intervention. For example, if the placebo group did not decline appreciably over a short period, it would be very difficult to demonstrate benefit of the experimental intervention. On the other hand, 18 months also was chosen as being short enough to ensure that an adequate number of patients would complete the trial.

Types of Interventional Study

Cross-over Trials

Senn defines the cross-over trial as one "in which subjects... are given a number of treatments with the object of studying differences between these treatments." [57] Double-blind cross-over trials involve administration of treatments in a sequence unknown to the patient or to the examiner. One of the treatments may be a placebo.

Cross-over trials enjoy a number of benefits. Patients serve as their own controls. This provides a savings in sample size compared with parallel group studies because the variability of differences within individual patients usually is substantially less than the variability of differences among patients. Fewer patients are required to generate the same total number of

observations; hence, patient recruitment is easier, and costs tend to be lower than in parallel group studies. On the other hand, the greater length of the study for each patient actually may hinder recruitment.

In an ideal cross-over experiment, the subject is in the same state at the start of each treatment period. Therapies for the short-term symptomatic improvement of a chronic disease (such as levodopa treatment of PD) are most suitable for this type of study because the underlying disease process is not expected to change significantly during the course of a brief trial.

Potentially serious drawbacks of the cross-over design include the difficulty of analysis in the event of withdrawals from the trial. If a patient serves as his or her own control, withdrawal at any point in the trial may render that patient's

460



Figure 1. Power curves for two-sided t -tests at the 5% level of significance for the difference in mean 18-month decline in total functional capacity (TFC) between a treated and an untreated (placebo) group of Huntington's disease patients. The expected decline in the untreated group is 1.0 +/- 1.5 units/year (mean +/- standard deviation) based on natural history. Curves are drawn for detecting differences in mean decline of 0.57, 0.50, and 0.46 TFC units/year between the two groups. The planned total sample size for the trial is 296 patients (148 patients per group).

461

data uninterpretable. Trials, therefore, must try to minimize the required length of subjects' participation.

Perhaps the most important potential drawback of cross-over designs, however, is that they are predicated on the assumption that neither passage of time itself, nor treatment order, affects the response to the currently administered treatment. [33] These treatment-by-period interactions or enduring carry-over effects may be extremely difficult to detect statistically and may hamper attempts to arrive at reasonable estimates of the actual overall treatment effects. To overcome the potentially confounding influence of treatment order on outcome, washout periods between treatments need to be long enough to eliminate residual drug effects. This need to lengthen a subject's participation in the study is at obvious odds with the desire to keep it brief. A reasonable compromise may be difficult to reach.

Parallel Group Trials

Unlike cross-over trials, parallel group controlled trials are less susceptible to the problem of variation in the disease over time. Because parallel designs do not involve using the patients as their own controls, withdrawals produce fewer problems in analysis, and carry-over effects are not an issue. In these trials, subjects are allocated to a given treatment arm and prospectively followed over time. A potential disadvantage of the parallel group design is that it may require more subjects and, therefore, added costs to implement.

One possible way of increasing the efficiency of parallel group trials is to study more than one intervention concurrently. For example, the DATATOP trial of selegiline HCl and tocopherol treatment in 800 early-stage PD patients incorporated a 2x2 factorial design. [45] Subjects were assigned randomly with equal allocation to one of four treatment groups: placebo, tocopherol alone, selegiline HCl alone, or the combination of selegiline HCl and tocopherol. Under certain assumptions, such a factorial design allows for the assessment of the effect of two interventions with the same sample size as would have been used in a single-intervention study. [11] Thus, DATATOP was much less costly than independent trials of selegiline HCl and tocopherol alone would have been.

The efficiency of the factorial design is sacrificed, however, if there is a negative statistical interaction (subadditivity) among the interventions being studied. This occurs if the efficacy for the combined treatment group is less than the sum of the efficacies in the groups in which each intervention is administered alone. For example, one drug may interfere with the absorption or distribution of the other drug, or the central pharmacodynamic effects of the two drugs may counteract each other. Subadditivity also may be induced by relatively poor compliance in the combined treatment group, or it simply may be

due to the measurement scale of the response variable. [2] In the DATATOP trial, there was a sound scientific rationale for believing that the effects of each agent were mediated through independent pharmacologic mechanisms. Also, the known safety profile of the two agents rendered compliance a minor issue. At the planning stages of a trial, the possibility of some subadditivity cannot be ruled out; therefore, it is wise to plan for a sample size sufficient to detect somewhat attenuated treatment effects in trials having a factorial design.

It should be noted, finally, that factorial designs often are planned for the sole purpose of examining interactions among treatments. Many subjects typically are required for such studies.

462

CONDUCTING CLINICAL TRIALS

The overarching mandate in the conduct of clinical trials is to ensure the safety of enrolled patients. The ancient admonition, *primum non nocere*--first of all, do no harm--demands that investigators adhere scrupulously to the constraints of a protocol once it has been approved by an institutional review board. Aspects of the protocol designed to maintain patient safety include specific inclusion and exclusion criteria, strict guidelines for suspension of study drug, reporting of adverse events, and emergency disclosure of treatment. The oversight of an independent safety monitoring committee is desirable in case the study needs to be terminated ahead of schedule due to accumulating evidence of a threat to patient safety. [4]

Many studies incorporate plans for interim analyses because strong evidence of the beneficial or harmful effects of a given treatment may disturb clinical equipoise and mandate the early termination of the trial. Such analyses should be performed by an independent group of clinical trial experts who are not involved in subject evaluation or the conduct of the study. Hall and Pledger [26] highlighted the possibility that interim analyses may indicate that a study is incapable of settling the question at hand, thereby rendering it futile to proceed. For example, such analyses may reveal that the degree of variability in the primary response variable is higher than had been anticipated and that the study's sample size, therefore, is inadequate to demonstrate a difference among treatment groups. Alternatively, interim analyses may indicate no significant differences among the groups despite a well-planned study. Investigators then may be confronted with the decision whether to continue the trial given that the magnitude of the treatment benefit is likely to be small at best. Factors favoring discontinuation of the study under these circumstances include the potential toxicity of the intervention and the costliness of continuing a trial that is not likely to show a benefit.

Organization, teamwork, and meticulous adherence to protocol are extremely important in the conduct of clinical trials. As the level of sophistication in our understanding of trial methodology has increased, so too has the degree of complexity involved in trial management.

The desire to detect small incremental improvements in the outcomes of patients has mandated the development of multicenter clinical trials. A relatively small impact on stroke reduction may have major public health implications, thereby justifying large trials in carotid stenosis or atrial fibrillation. Multicenter trials also are necessary for the investigation of therapies in relatively rare disorders when a single center may have difficulty enrolling adequate numbers of patients. Regardless of disease prevalence, multicenter trials increase the generalizability of conclusions by broadening the sample of patients studied.

Several articles in a recent supplement of *Controlled Clinical Trials* have focused on the fundamental issues of data collection, storage, and processing in multicenter trials. [24] [29] [35] [37] In addition to investigators and coordinators, a sizeable team of biostatisticians, administrators, monitors, programmers, and clerical staff is necessary to meet the demands of a modern multicenter trial.

The coordination center for a clinical trial has many responsibilities, including production of study protocols, oversight of data collection, data processing, reporting of adverse experiences, maintenance of study files, and administration of grant funds. As the central clearing-house for the trial, the coordination center provides liaison with individual sites. The electronic transfer of data enables the rapid distribution of important information regarding enrollment and adverse events. Database organization and error checking are central responsibilities of

coordination center staff. A well-organized and well-run coordination center is a key component in the successful planning and implementation of a multicenter clinical trial.

INTERPRETING CLINICAL TRIALS

In order to be worthwhile, clinical trials must generate results that extend beyond the narrow boundaries of the patient sample studied. The interpretation of clinical trials requires a clear sense of both the sample of patients studied and the population from which this sample was drawn. The function of inclusion and exclusion criteria in trial protocols is to define clearly the population about whom conclusions are to be made. Narrow inclusion criteria may maximize the ability of the investigators to demonstrate an effect of the intervention on the study sample but may limit the applicability of conclusions to the broader population. It also must be kept in mind that patients who consent to participate in trials often differ in important respects from those who are eligible but decline participation.

Davis has pointed out that "several pieces of information are necessary to determine the extent of extrapolation or generalization warranted in a specific clinical trial." [19] He stressed the need for integration of information from basic laboratory studies, observational studies, and other clinical trials.

The choice of outcome variable is extremely important in the design and interpretation of clinical trials. The ideal variable would be readily measured, clinically meaningful, and biologically relevant. In most neurologic disorders, however, it is rare for a single variable to meet all of these constraints; several outcome variables are measured commonly in clinical trials. It is, nonetheless, desirable for investigators to identify the primary outcome variable before initiating a trial. This identification enables rational calculation of necessary sample size, compels investigators to organize their approach, and leads to more efficient use of resources. Prior specification of the primary outcome measure also precludes the problems associated with multiple statistical testing and post-hoc data dredging.

The primary outcome variable often is the most clinically meaningful or most objectively measurable one (e.g., mortality). Ideally, the outcome should be defined clearly and should have well-established reliability and validity. Objectivity in and of itself, however, is not necessarily the only criterion to be met. In recognition of this fact, many trials have incorporated global assessments of efficacy in which subjects and investigators give an overall impression in terms of "better," "unchanged," or "worse." These subjective ratings perhaps are better suited to address quality-of-life issues.

The question of what represents a clinically meaningful improvement can be difficult to settle. A recent trial of riluzole in amyotrophic lateral sclerosis indicated that this drug prolonged life (primary outcome) but failed to demonstrate improvement in motor function (secondary outcome). [9] Similarly, the relationship between functional decline and radiographic evidence of white matter lesions has been debated in trials of multiple sclerosis. [49]

The growing role of meta-analyses in the literature supports the notion that the sum of information about a given intervention must be taken into account when deciding whether that intervention is truly safe and effective. Meta-analyses involve the pooling of results from several clinical trials in an attempt to extrapolate about the effect of an intervention. Examples include recent articles on dexamethasone in the treatment of bacterial meningitis, [25] antithrombotic

therapy in atrial fibrillation, [3] and ergoloid mesylates in the treatment of dementia. [55] Ultimately, the interpretation of clinical trials must be predicated on the care with which they are designed and conducted.

References

1. Armitage P, Berry G: The analysis of covariance. *In* Statistical Methods in Medical Research, ed 3. Oxford, England, Blackwell Scientific, 1994, p 301

2. Armitage P, Berry G: Comparison of several groups. *In* *Statistical Methods in Medical Research*, ed 3. Oxford, England, Blackwell Scientific, 1994, p 226
3. Atrial Fibrillation Investigators: Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation: Analysis of pooled data from five randomized controlled trials. *Arch Intern Med* 154:1449, 1994
4. Barnett HJM, Sackett DL: Monitoring clinical trials. *Neurology* 43:2437, 1993
5. Bensimon G, Lacomblez L, Meininger V, et al: A controlled trial of riluzole in amyotrophic lateral sclerosis. *N Engl J Med* 330:585, 1994
6. Bourgeois B, Leppik IE, Sackellares JC, et al: Felbamate: A double-blind controlled trial in patients undergoing presurgical evaluation of partial seizures. *Neurology* 43:693, 1993
7. Brittain E, Wittes J: Factorial designs in clinical trials: The effects of non-compliance and subadditivity. *Stat Med* 8:161, 1989
8. Brooke MH, Fenichel GM, Griggs RC, et al: Clinical investigation in Duchenne dystrophy: 2. Determination of the "power" of therapeutic trials based on the natural history. *Muscle Nerve* 6:91, 1983
9. Bull JP: The historical development of clinical therapeutic trials. *Journal of Chronic Disease* 10:218, 1959
10. Butterfield PG, Valanis BG, Spencer PS, et al: Environmental antecedents of young-onset Parkinson's disease. *Neurology* 43:1150, 1993
11. Byar DP, Piantadosi S: Factorial designs for randomized clinical trials. *Cancer Treatment Reports* 69:1055, 1985
12. Carlsson A: The occurrence, distribution, and physiological role of catecholamines in the nervous system. *Pharmacol Rev* 11:490, 1959
13. Cereghino JJ: Clinical trial design for antiepileptic drugs. *Ann Neurol* 32:393, 1992
14. Chalmers TC: The control of bias in clinical trials. *In* Shapiro SH, Louis TA (eds): *Clinical Trials: Issues and Approaches*. (Statistics: Textbooks and Monographs, vol 46.) New York, Marcel Dekker, 1983
15. Charcot JM: Lectures on the diseases of the nervous system. London, The New Sydenham Society, 1877, p 279
16. Colton T: Medical surveys. *In* *Statistics in Medicine*. Boston, Little Brown, 1974
17. Cotzias GC, VanWoert MH, Schiffer LM: Aromatic amino acids and modification of parkinsonism. *N Engl J Med* 280:337, 1967
18. Cunningham RW, Harned BK, Clark MC, et al: The pharmacology of 3-(N-piperidyl)-1-phenyl-1-cyclohexyl-1-propanol HCl (Artane) and related compounds: New antispasmodic agents. *J Pharmacol Exp Ther* 96:151, 1949
19. Davis CE: Generalizing from clinical trials. *Control Clin Trials* 15:11, 1994
20. Dawson-Saunders B, Trapp RG: Study designs in medical research. *In* *Basic and Clinical Biostatistics*, Norwalk, CT, Appleton & Lange, 1994, p 6
21. Ellenberg JH: Clinical trials. *Neurol Clin* 8:20, 1990
22. Fehling C: Treatment of Parkinson's syndrome with L-dopa, a double blind study. *Acta Neurol Scand* 1:180, 1966
23. Freedman B: Equipoise and the ethics of clinical research. *N Engl J Med* 317:141, 1987
24. Gassman JJ, Owen WW, Kuntz TE, et al: Data quality assurance, monitoring and reporting. *Control Clin Trials* 16:104S, 1995
25. Geiman BJ, Smith AL: Dexamethasone and bacterial meningitis: A meta-analysis of randomized controlled trials. *West J Med* 157:27, 1992
26. Hall D, Pledger G: Early termination of clinical trials. *Journal of Clinical Research and Drug Development* 2:83, 1988

27. Hellman S, Hellman DS: Of mice but not men: Problems of the randomized clinical trial. *N Engl J Med* 324:1585, 1991
28. Hibberd PL: Use and misuse of statistics for epidemiological studies of multiple sclerosis. *Ann Neurol* 36(suppl 2):S218, 1994
29. Hosking JD, Newhouse MM, Bagniewska A, et al: Data collection and transcription. *Control Clin Trials* 16:66S, 1995
30. Huntington G: On chorea. *Medical and Surgical Reporter* 26:320, 1872
31. Inzitari D, DiCarlo A, Mascalchi M, et al: The cardiovascular outcome of patients with motor impairment and extensive leukoaraiosis. *Arch Neurol* 52:687, 1995
32. Lebacqz K: Ethical aspects of clinical trials. *In* Shapiro SH, Louis TA (eds): *Clinical Trials: Issues and Approaches (Statistics: Textbooks and Monographs, vol 46.)* New York, Marcel Dekker, 1983, p 94
33. Louis TA, Lavori PW, Bailar JC III, et al: Crossover and self-controlled designs in clinical research. *In* Bailar JC III, Mosteller F (eds): *Medical Uses of Statistics.* Waltham, MA, New England Journal of Medicine Books, 1986, p 81
34. Mark MH, Duvoisin RC: The history of the medical therapy of Parkinson's disease. *In* Koller WC, Paulson G (eds): *Therapy of Parkinson's Disease: Second Edition Revised and Expanded. (Neurological Disease and Therapy, vol 34.)* New York, Marcel Dekker, 1995, p 8
35. McBride R, Singer SW: Interim reports, participant closeout, and study archives. *Control Clin Trials* 16:137S, 1995
36. McDowell FH, Lee JE, Sweet R, et al: The treatment of Parkinson's disease with dihydroxyphenylalanine. *Ann Intern Med* 72:19, 1970
37. McFadden ET, LoPresti F, Bailey LR, et al: Approaches to data management. *Control Clin Trials* 16:30S, 1995
38. Mendell JR, Province MA, Moxley RT III, et al: Clinical investigation of Duchenne muscular dystrophy: A methodology for therapeutic trials based on natural history controls. *Arch Neurol* 44:808, 1987
39. Montagu KA: Catechol compounds in rat tissues and in brains of different animals. *Nature* 180:244, 1957
40. North American Symptomatic Carotid Endarterectomy Trial Collaborators: Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade stenosis. *N Engl J Med* 325:445, 1991
41. Noseworthy JH, Ebers GC, Vandervoort MK, et al: The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology* 44:16, 1994
42. Parkinson JD: *An Essay on The Shaking Palsy.* London, Sherwood, Neely & Jones, 1817
43. Parkinson Study Group: A controlled trial of lazabemide (R019-6327) in untreated Parkinson's disease. *Ann Neurol* 33:350, 1993
44. Parkinson Study Group: A controlled trial of lazabemide (R019-6327) in levodopa treated Parkinson's disease. *Arch Neurol* 51:342, 1994
45. Parkinson Study Group: DATATOP: A multicenter controlled clinical trial in early Parkinson's disease. *Arch Neurol* 46:1052, 1989
46. Parkinson Study Group: Effect of deprenyl on the progression of disability in early Parkinson's disease. *N Engl J Med* 321:1364, 1989
47. Parkinson Study Group: Effects of tocopherol and deprenyl on the progression of disability in early Parkinson's disease. *N Engl J Med* 328:176, 1993
48. Passamani E: Clinical trials: Are they ethical? *N Engl J Med* 324:1589, 1991
49. Paty DW, Li DKB, Oger JJ, et al: Magnetic resonance imaging in the evaluation of clinical trials in multiple sclerosis. *Ann Neurol* 36:S95, 1994
50. Pocock SJ: Blinding and placebos. *In* *Clinical Trials: A Practical Approach.* Chichester, John Wiley and Sons, 1983, p 90
51. Pocock SJ: The combination of randomized and historical controls in clinical trials. *Journal of Chronic Disease* 29:175, 1976

52. Pocock SJ: The historical development of clinical trials. *In* Clinical Trials: A Practical Approach. Chichester, England, John Wiley and Sons, 1983, p 26
53. Pocock SJ: The justification for randomized controlled trials. *In* Clinical Trials: A Practical Approach. Chichester, England, John Wiley and Sons, 1983
-

466

54. Pradas J, Finison L, Andres PL, et al: The natural history of amyotrophic lateral sclerosis and the use of natural history controls in therapeutic trials. *Neurology* 43:751, 1993
55. Schneider LS, Olin JT: Overview of clinical trials of hydergine in dementia. *Arch Neurol* 51:787, 1994
56. Semchuk KM, Love EJ, Lee RG: Parkinson's disease and exposure to agricultural work and pesticide chemicals. *Neurology* 42:1328, 1992
57. Senn S: *Cross-over Trials in Clinical Research*. Chichester, England, John Wiley and Sons, 1993, p 1
58. Shoulson I, Fahn S: Huntington's disease: Clinical care and evaluation. *Neurology* 29:1, 1979
59. Standards of Reporting Trials Group: A proposal for structured reporting of randomized controlled trials. *JAMA* 272:1926, 1994
60. Stern M, Dulaney E, Gruber SB, et al: The epidemiology of Parkinson's disease: A case-control study of young-onset and old-onset patients. *Arch Neurol* 48:903, 1991
61. Yahr MD, Duvoisin RC, Schear MJ, et al: Treatment of parkinsonism with levodopa. *Arch Neurol* 21:343, 1969

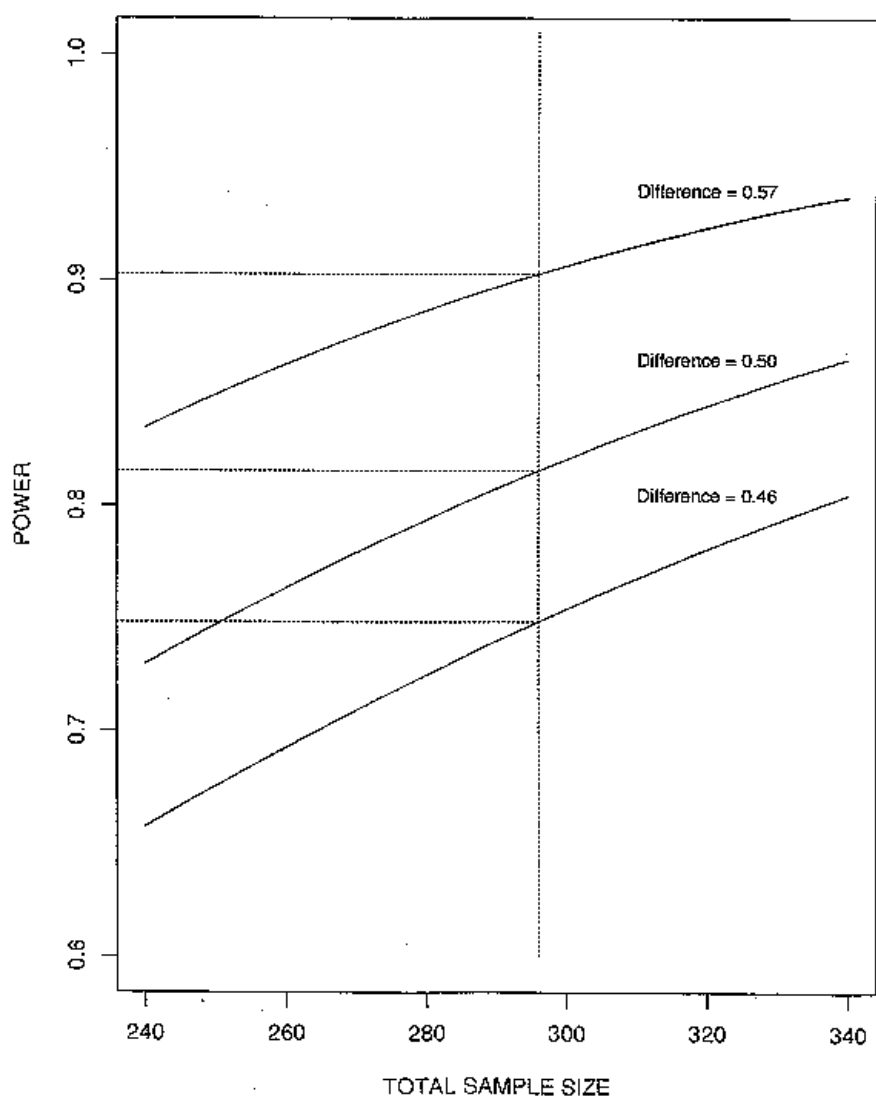


Figure 1. Power curves for two-sided *t*-tests at the 5% level of significance for the difference in mean 18-month decline in total functional capacity (TFC) between a treated and an untreated (placebo) group of Huntington's disease patients. The expected decline in the untreated group is 1.0 ± 1.5 units/year (mean \pm standard deviation) based on natural history. Curves are drawn for detecting differences in mean decline of 0.57, 0.50, and 0.46 TFC units/year between the two groups. The planned total sample size for the trial is 296 patients (148 patients per group).