Review

Guidelines for the conduct of clinical trials for spinal cord injury as developed by the ICCP panel: clinical trial design

D Lammertse*[1,2], MH Tuszynski[3,4], JD Steeves[5], A Curt[5], JW Fawcett[6], C Rask[7], JF Ditunno[8], MG Fehlings[9], JD Guest[10], PH Ellaway[11], N Kleitman[12], AR Blight[13], BH Dobkin[14], R Grossman[15], H Katoh[16], A Privat[17] and M Kalichman[18]

[1]Craig Hospital, Englewood, CO, USA; [2]Department of Physical Medicine and Rehabilitation, University of Colorado at Denver and Health Sciences Center, Denver, CO, USA; [3]Center for Neural Repair, Department of Neurosciences, University of California – San Diego, La Jolla, CA, USA; [4]Department of Neurology, Veterans Administration Medical Center, La Jolla, CA, USA; [5]ICORD, University of British Columbia and Vancouver Coastal Health Research Institute, Vancouver, British Columbia, Canada; [6]Cambridge University Centre for Brain Repair, Cambridge, UK; [7]Institute for OneWorld Health, San Francisco, CA, USA; [8]Department of Rehabilitation Medicine, Jefferson Medical College of Thomas Jefferson University, Philadelphia, PA, USA; [9]Department of Surgery, Division of Neurosurgery, University of Toronto, Toronto, Ontario, Canada; [10]University of Miami, Department of Neurosurgery, Miami, FL, USA; [11]Department of Movement & Balance, Division of Neuroscience & Mental Health, Imperial College London, Charing Cross Campus, London, UK; [12]National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA; [13]Acorda Therapeutics Inc., Hawthorne, NY, USA; [14]Department of Neurology, University of California at Los Angeles, Los Angeles, CA, USA; [15]Methodist Neurological Institute, Houston, TX, USA; [16]Keio University School of Medicine, Tokyo, Japan; [17]Institut des Neurosciences – CHU St Eloi, INSERM U-583, Montpellier cedex, France; [18]Research Ethics Program, Department of Pathology, University of California at San Diego, La Jolla, CA, USA

The International Campaign for Cures of Spinal Cord Injury Paralysis established a panel tasked with reviewing the methodology for clinical trials for spinal cord injury (SCI), and making recommendations on the conduct of future trials. This is the fourth of four papers. Here, we examine the phases of a clinical trial program, the elements, types, and protocols for valid clinical trial design. The most rigorous and valid SCI clinical trial would be a prospective double-blind randomized control trial utilizing appropriate placebo control subjects. However, in specific situations, it is recognized that other trial procedures may have to be considered. We review the strengths and limitations of the various types of clinical trials with specific reference to SCI. It is imperative that the design and conduct of SCI clinical trials should meet appropriate standards of scientific inquiry to insure that meaningful conclusions about efficacy and safety can be achieved and that the interests of trial subjects are protected. We propose these clinical trials guidelines for use by the SCI clinical research community.
Spinal Cord (2007) **45**, 232–242. doi:10.1038/sj.sc.3102010; published online 19 December 2006

Introduction

This is the fourth paper of a series of four, reporting the deliberations of a panel sponsored by the International Campaign for Cures of Spinal Cord Injury Paralysis (ICCP), an affiliation of 'not for profit' organizations, which has an interest in facilitating the translation of valid treatments for spinal cord injury (SCI) paralysis. The four papers address considerations relevant to the planning and design of therapeutic trials in SCI. Their subjects are:

1. Natural history of SCI, degree and time course of spontaneous recovery and statistical power needed to achieve a valid outcome;

*Correspondence: D Lammertse, Craig Hospital, 3425 S. Clarkson Street, Englewood, Colorado, 80113, USA

2. appropriate clinical outcome measures for different clinical phases and targets;
3. patient selection criteria (inclusion/exclusion), confounding variables, and ethics;
4. trial design, statistical analysis, and organization of multicenter trials.

This fourth paper from the ICCP Clinical Guidelines Panel addresses the topic of the overall design of a clinical trial, including multicenter studies.

Review of the phases of clinical trials

All the phases of a clinical development program (described below) are usually necessary for the appropriate regulatory body to grant the approval of the therapeutic for clinical use, and each phase may involve a number of successive or parallel trials. Most of the principles of standard drug trials apply to potential SCI therapeutics, but there are additional points unique to SCI trials that should also be considered, as pointed out below.

Phase 1 – safety
Phase 1 begins with the first administration of the therapeutic intervention (eg, experimental drug or cellular transplant) to a human subject, and explores a number of aspects on the safety and interaction between treatment and subject that can be examined in an 'open' unblinded way. The first human study is usually based on extensive preclinical safety evaluations, and has a built-in margin of safety between the highest doses and durations of treatment explored in animal studies and the initial human protocol. Evaluation of safety is an important aspect of all phases of clinical development, but it is particularly prominent in Phase 1 (or 'pilot') trials that can expose the most common adverse events (side effects or complications) of any intervention. There is often an element of upward dose-exploration in Phase 1, especially in cases where there is (or is suspected to be) a narrow therapeutic index (the separation between dose ranges that are effective and those that are tolerable and safe). In these cases, it may be important to identify a maximum tolerated dose, whereas later phase studies may be able to identify a minimum effective dose. In the special case of biologic therapeutics, this is sometimes the maximum feasible dose, and no maximally tolerated dose can be identified.

Another important aspect of a Phase 1 drug studies is measurement of the pharmacokinetics, and potentially the pharmacodynamics of the therapeutic. Pharmacokinetics is the study of the temporal and spatial fate of drugs in the body, with emphasis on the time required for absorption, distribution within body tissues, the mode and extent of metabolism, and the method of excretion. Pharmacodynamics is the study of the biochemical and physiological effects of drugs and the mechanisms of drug action, and the relationship between drug concentration and effect. It is often summarily stated that pharmacodynamics is the study of what a drug does to the body, whereas pharmacokinetics is the study of what the body does to a drug. Pharmacokinetics is typically studied early in the process of development, whereas pharmacodynamics may not be clear until after the effects of the compound are more fully understood. In the case of cellular treatments, equivalent studies of the fate of implanted cells or tissues will be important to pursue, but may be much more difficult to implement.

In the clinical setting, the potential therapeutic results of a treatment depend on both pharmacokinetics and pharmacodynamics. For example, a patient may have excessive toxicity at a particular dose for several reasons. If the participant's pharmacokinetics are different from those of the typical patient, there may be decreased total body clearance, resulting in higher than expected levels of active drug within the body (ie abnormal pharmacokinetics). Alternatively, the subject might be more sensitive to the experimental drug (potentially producing altered pharmacodynamics). Thus, an evaluation utilizing multiple subjects across a range of different doses is helpful in defining an optimal therapeutic regimen for subsequent clinical trial phases.

Phase 1 trials usually involve a small sample size (the size of this may vary depending on whether a dose escalation of a pharmacological compound is required). Phase 1 trials may, but more often do not, include randomized control subjects, and they are usually carried out in an unblinded fashion where both the participants and investigators know what drug is being tested and what dosages are being used (sometimes referred to as 'open label'). Phase 1 studies of noninvasive or minimally invasive treatments are often undertaken in healthy volunteer subjects. As many of the potential SCI therapies are likely to be invasive, Phase 1 trials would be expected to involve subjects with SCI. As a result, there is an opportunity to undertake a preliminary evaluation of the possible therapeutic benefit of the experimental treatment. This can be useful for identifying appropriate clinical end points and outcome measures for subsequent trial phases. Choice of the appropriate outcome measures is essential to determine accurately both safety and efficacy.[1]

Phase 2 – therapeutic exploratory
In Phase 2 trials, the primary objective shifts to the exploration of potential therapeutic effect size and variability in patients compared to a control group, with a determination of the most appropriate outcome measures with which to detect potential therapeutic effects. Thus, a Phase 2 trial is designed to demonstrate the 'activity' of an intervention: that is, to demonstrate that the intervention is associated with a positive change in relevant outcome variables, often using less stringent statistical criteria than Phase 3 trials, or using surrogate markers as outcome measures. Nonetheless, analysis of outcomes in early Phase 2 trials can provide important guidance for refinements in the treatment regimen and

outcome measurement for subsequent Phase 2 and Phase 3 trials.

There are a number of protocol designs for Phase 2 trials (see below), but all trials at this stage should include control subjects and some form of 'blinded' assessment, where the person undertaking the outcome measurement and/or evaluating the outcome data does not know the treatment or control group assignment of the subject. The preferred Phase 2 design would be a randomized control trial (RCT), where each participant is recruited prospectively and randomly assigned to either the experimental or control arm of the study, and where the investigators and, if at all possible, the participants are blinded to which study arm each participant has been assigned.

Another common characteristic of Phase 2 trials is the use of relatively narrow inclusion criteria to ensure a more homogeneous study cohort. For example, it may not be optimal in a Phase 2 trial to simultaneously compare data from motor complete (ASIA A and B) with motor incomplete (ASIA C and D) subjects. To avoid comparing 'apples with oranges', many Phase 2 trials have different study arms or cohorts of subjects, which are distinctive from other groups.

A Phase 2 study provides further opportunity to further refine the optimal dose, timing, and treatment regimen (eg, concomitant interventions, drug infusion or cellular transplant location, and other potential confounding variables) for the more definitive Phase 3 trial. Even though most Phase 2 trials declare a primary clinical end point and outcome threshold, they should also evaluate a number of different clinical endpoints (secondary outcomes) to guide the selection of the most definitive Phase 3 primary outcome. It is not uncommon to undertake more than one Phase 2 trial to explore other target populations that might receive benefit from the therapeutic agent. Although a Phase 2 study usually involves a fairly limited number of subjects, it is designed to identify whether a therapeutic effect is likely to be present, to gather more evidence of the intervention's safety, and refine the parameters for a more comprehensive Phase 3 trial.

In rare cases, a Phase 1 trial may be combined with a Phase 2 trial design, in which a larger number of subjects are recruited with the view of combining a safety study with a trial designed to obtain preliminary data on possible therapeutic efficacy. This requires the recruitment of subjects in which the natural history of neurological recovery (or lack thereof) is predictable and not so variable as to hide treatment effects. Depending on the robustness of the data, this may allow, evolution from a Phase 1/2 trial directly to a Phase 3 RCT, though this situation is uncommon, and not likely to be seen in a treatment for a condition as difficult to address as SCI.

Phase 3 – therapeutic confirmatory
Phase 3 clinical trials are generally the definitive clinical trial phase, and are typically undertaken as a randomized, controlled trial. The objective is to confirm the preliminary evidence obtained at the Phase 2 stage with a statistically significant clinical benefit of the therapeutic in a larger group of subjects across multiple study centers. Given that the Phase 2 trial may have been conducted on a well-defined subset of patients with SCI, it is also possible to consider including a broader spectrum of subjects in a Phase 3 study. For example, if the Phase 2 study was only conducted on acute SCI patients who are motor-complete (ie ASIA A and B), a Phase 3 might now include patients with motor incomplete SCI (ie ASIA C and D) or other types of incomplete SCI, such as central cord syndrome, or cauda equina injury. However, such a broadening of patient selection could increase the risk of failure in the larger study, as the treatment may not prove efficacious in all forms of SCI, or may pose a previously unforeseen risk when tested in an expanded range of subjects.

As the patient population under investigation is expanded to include a more heterogeneous group of subjects, appropriate sizing of the trial and consideration of stratification strategies become critically important (cf Steeves et al[1]). For this reason, it is best to design a Phase 3 protocol based closely on the design features of previous, smaller Phase 2 studies that allow a relevant power analysis to be made. A relevant power analysis can only be based on experience with the same study parameters. Any change in the trial procedures or the characteristics of the trial participants will adversely affect the ability to predict the behavior of the new study from that of the old.

Depending on the strength of the clinical benefit provided by the therapeutic intervention, and careful analysis of existing data, a Phase 3 trial might also be expanded to include subjects with injuries in a broader interval of time-after-injury (eg, the study of an acute intervention might be expanded to include subacute injury subjects). Such broadening of inclusion criteria at the stage of Phase 3 investigation should be supported by preclinical data, indicating efficacy at corresponding intervention time frames, and preceded by examination in a separate Phase 2 study, where dose–response relationships could be adjusted to the specific pharmacokinetics or pharmacodynamics of the new, expanded patient population. Only then would it be possible to power the Phase 3 study appropriately.

If the Phase 3 investigation concludes with the valid demonstration of a statistically significant clinical benefit from the therapeutic and an acceptable adverse event profile, an application is usually made to the appropriate regulatory body for approval to market the treatment. Some jurisdictions, particularly the United States, prefer that a second confirmatory Phase 3 trial be completed before approval is granted.

Phase 4 – therapeutic use
Phase 4 begins with marketing approval and introduction of the therapeutic intervention for clinical use. It

includes ongoing surveillance related to therapeutic safety, including possible drug interactions and contra-indications, continued optimization of dose–response relationships and therapeutic delivery regimens, as well as studies to delineate additional information on the intervention's risks, benefits, and optimal use.

Clinical trial design

Good clinical trial design includes the elimination or minimization of extraneous variables within the trial design and an accounting of the remaining variables in all reports (also see Tuszynski et al[3]). The outcome measures should be specific to the behavior or function being assessed,[1] and enable functional clinical benefit to be demonstrated with statistical significance in an adequately powered pivotal study (also see Fawcett et al[2]).

It must be remembered that it is relatively easy to get positive responses from subjects in a trial when they know they have been treated with an experimental therapy, and when they expect or hope for a benefit (the so-called 'placebo effect'). Thus, trials with randomized controls and blinded assessments are necessary standards to remove investigator and subject bias. Finally, all trial results, whether positive or negative, should be registered with regulatory agencies and submitted for publication.

Bias is defined here as the systematic tendency of any factors associated with the design, conduct, analysis, and interpretation of the results of clinical trials to make the estimate of a treatment effect (therapeutic benefit) deviate from its true value (see below). Each outcome measure must accurately and sensitively track any changes in the behavior or function being evaluated. An outcome measure should have both precision (ie consistency) and robustness, which means the overall findings are not significantly influenced by slight variations in treatment regimens, assessment procedures, or data analysis.

Numerous guidelines for the general conduct of any, and all, clinical trials have been developed, and readers are encouraged to make themselves familiar with these teachings, especially those developed by the International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use (available at www.ich.org). When considering the design and development of a SCI clinical trial, the following ICH documents may be of most interest: (available at www.ich.org/LOB/media/)

E3, Structure and Content of Clinical Study Reports
E6, Good Clinical Practice: Consolidated Guidance
E8, General Considerations for Clinical Trials
E9, Statistical Principles for Clinical Trials
E10, Choice of Control Group and Related Issues in Clinical Trials

The United States FDA website also provides its guidelines and those of the ICH at its website (www.fda/cder/guidance/index.htm).

Clinical trial protocol designs

There are numerous trial configurations and each has its particular strengths and limitations. An important concern for all clinical trials is the potential for bias, however unintentional, to influence the interpretation of clinical outcomes. There are varying degrees of blinding, starting with 'Open Label', wherein the identity of the treatment is known to both the investigators and participants. This should generally be reserved for Phase 1 (safety) trials only. 'Open Label' protocols have been used in the study of both pharmacological and surgical SCI interventions in Phase 1 trials.[4–8]

The next level is a 'Single Blind' study where either the clinical investigator or the subject, but not both, are blinded. For SCI trials where a surgical intervention is part of the experimental protocol, it may be necessary for the surgeon to know what is being undertaken in that subject. However, it is preferred that the patient remains blinded to the treatment received (for both experimental and control groups), although this is not always possible. Nevertheless, independent outcome assessors/examiners should remain blinded to the treatment provided. This may require monitoring to assure that a subject does not disclose to the assessor the treatment arm to which they have been assigned. Ethical or legal difficulties may interfere with the use of blinding when it entails sham operative procedures. Nonetheless, sham surgical trials have been implemented in neurological disorders in recent years, and they should be considered in SCI trials as well.[9,10] In any event, outcome assessments should be blinded using techniques such as identical bandaging of the overlying skin during assessments by independent examiners. Single blinding of a primary outcome measurement has been utilized in recent Phase 2 randomized, controlled trials of autologous macrophages in the treatment of subacute SCI.[7]

Finally, in an optimal 'Double Blind' design, neither the participating trial subject nor the investigators or sponsor staff are aware of the treatment received during the trial.[3] Ideal blinding would ensure that the treatments cannot be distinguished by subjective experience, appearance, timing, and delivery method by any of the subjects, investigators, research staff, or clinical staff. This should be maintained throughout the conduct of the entire trial from determination of eligibility through evaluation of all endpoints, and requires full compliance of the subject. Double Blind design has been used in a number of pharmacological trials in SCI, including investigations of methylprednisolone and GM-1 ganglioside in acute injury[11–15] and 4-aminopyridine in chronic spinal injury,[16,17] and in more recent surgical trials for Parkinson's disease.[9,10]

Randomization in the assignment of trial participants to the different study arms (groups), including a placebo (or standard of care) control group, is done to reduce

bias, and introduces a deliberate element of chance into the assignment of treatments. This provides a sound statistical basis for the quantitative evaluation of the evidence relating to treatment effects. Randomization tends to produce treatment groups where the distribution of prognostic factors (independent variables), known and unknown, are similar and representative of the overall patient population. Randomization should only be performed when eligibility of the subject for inclusion into the trial has been confirmed.[3] Randomization schemes for multicenter trials should be centrally organized and 'blocked' by center (ie the randomization scheme is separately applied to each center's group of subjects rather than to all subjects in the trial) in such a way that randomization occurs adequately at the individual center level as well as across the study as a whole.

Trial design configuration

Later in this review, we will discuss the relative merits of 'Frequentist' study designs (prevalent in current clinical trials) *versus* the emerging 'Bayesian' statistical trial designs, and how they influence the recruitment and assignment of participants to a study arm as well as the conduct and duration of a study. The most prevalent trial design configurations are provided below.

Parallel group design is the most common clinical trial design for pivotal Phase 3 trials. Subjects are randomized (often in equal numbers) to one or more treatment arms, each testing a different treatment or combination of treatments. The treatments might include the investigational product at one or more doses, and one or more control conditions such as a placebo (eg, the NASCIS 2 trial,[12]) and/or an active comparator (eg, the comparison of high- and low-dose methylprednisolone in the first NASCIS trial,[11]). A current treatment may have to be present in both the active and the control arms of the study, such as methylprednisolone in the multi-center GM-1 trial.[15] Assumptions underlying the parallel group design are less complex and more robust than those of other designs.

Crossover designs consist of subject randomization to a sequence of two or more treatments (eg, placebo control and experimental therapeutic). Hence, the subject acts as his or her own control for treatment comparisons. This approach has been used in the evaluation of 4-aminopyridine in chronic injury (Potter *et al*.[16]). To make valid assessments of functional efficacy for a treatment when subjects act as their own control, the clinical outcome measure must have a very stable (unchanging) baseline before application of the experimental treatment and the subsequent evaluations. Because the functional capacities of a person with acute or sub-acute SCI can vary dramatically over a short period of time, this type of design should be restricted to studies of chronic SCI, where the functional capacity to be assessed is expected to be stable (see Fawcett *et al*[2]). The relevant effects of treatment should develop fully

within the treatment period, and reverse following removal of treatment.

One important concern of a crossover design is the possibility of residual effects (carryover influence) of the experimental or placebo control treatment, which can influence the outcome after the subject has crossed over to the opposite treatment arm. Thus, the 'washout' time period between treatment arms should be sufficiently long to allow the complete reversibility of any treatment effect. However, an advantage of the crossover design is a reduction in the number of subjects or assessments needed to achieve a specific statistical power.

Group Sequential Design is used to facilitate an interim analysis, usually by an independent Data Safety Monitoring Board (DSMB) or Independent Data Monitoring Committee (IDMC). It involves any statistically valid analysis (eg, Bayesian statistics) intended to compare treatment arms with respect to safety or efficacy at any time before the proposed formal completion of the trial. Such an approach can be used to identify and discontinue an unsafe trial more quickly, stop a trial when the treatment has quickly demonstrated dramatic benefit, or identify an ineffective treatment dose with the subsequent randomization of subjects to a potentially more effective treatment arm. Utilization of such a design requires a declaration, at the beginning of a trial, that an interim analysis will be undertaken, and the application of prospectively defined distinct criteria (thresholds) for early termination of the study for safety or efficacy reasons.

Factorial Designs involve the testing of two or more treatments simultaneously for possible synergistic or antagonistic effects (combination treatments). In the simplest 2×2 form, subjects are allocated to one of four groups: A alone, B alone, both A and B, or neither A nor B. In 'Add-on' Trials, the test treatment and placebo are added to a common standard therapy.

Concerning to Factorial study design, there is an ongoing discussion in the community of spinal cord scientists and clinicians regarding the testing of combination therapies in clinical trials for SCI. There is the possibility that a combination of treatments could be tested in a clinical trial without including separate study arms that separately test only one element of the full combination. Clear preclinical animal data, would be required to support the feasibility of such a trial, and they should indicate that the combination therapy is effective, whereas individual components of the therapy administered in isolation are not. Safety and toxicity data would also be required to indicate that the individual components of the therapy are safe, when administered in isolation as well as when they are combined.

Types of trials

A trial to establish *Superiority* has the primary objective of demonstrating whether the investigational treatment has superior clinical benefit relative to a placebo or a comparative active therapy. In a superiority trial,

demonstration of a dose–response relationship between a therapeutic and a clinical measure is suggestive of efficacy. Trials of novel interventions for neurological improvement in SCI will generally consist of superiority trials.

If a treatment has been suggested in a previous clinical trial to be modestly efficacious, then a controlled trial of a novel therapeutic may need to incorporate a treatment arm that compares the outcome to the previous (modest) therapeutic. Indeed, depending on the efficacy of the previous (modest) therapeutic, withholding it from any treatment group, including that which receives the novel therapeutic, is an ethical challenge. A further caveat includes the possibility that the experimental treatment could interact adversely with the previous (conventional) therapeutic. This was an issue in the Sygen multicenter trial in which methylprednisolone was given to all the subjects, but had to be completed before initiation of study medication because of concerns based on pre-clinical data that concomitant administration of methylprednisolone and Sygen could negate the therapeutic effect of the investigational drug.[14,15] For this reason, the appropriateness of using a placebo control *versus* an active control should be carefully considered and based on prior knowledge of any adverse events, as well as pre-clinical examinations for undesirable drug interactions.

A trial to demonstrate *Equivalence* typically involves an attempt to demonstrate that a generic form of a drug has comparable efficacy to an approved drug (for which patent protection is ending). These are usually smaller trials, that rely on the demonstration of an equivalent chemical formulation of the drug in question, and on the previous clinical experience with that drug. Generic formulations can significantly reduce the cost of a treatment.

A *Non-inferiority* trial is a variation on the theme of demonstrating equivalence. It consists of an active control trial designed to show that the efficacy of a new investigational drug is no worse than that of the active comparator by a certain pre-specified margin (eg, 10%), and that the new drug is potentially safer. It is often used by competitors wishing to introduce their product to compete with an established drug. The issue of assay sensitivity is important in either an Equivalency or Non-inferiority trial, as the comparative 'new' treatment must be tested in the same manner (eg, dose, timing, regimen, primary outcomes) as used to establish efficacy of the original treatment (now the comparative control). Until approved therapies for neurological improvement in SCI are available, equivalence and non-inferiority trials will only be appropriate considerations for comparing of investigational treatments with approved *symptomatic* treatments (eg, antispasticity medications).

Multicenter SCI clinical trials

Given the relatively low incidence of SCI on an annual basis (often less than 40 cases per million population),

SCI is appropriately classified as a rare disorder. Thus, even with broad inclusion and minimal exclusion criteria, it is often difficult to recruit sufficient participants for any trial phase within one study center. The only acceptable means of accruing sufficient numbers of subjects to satisfy the trial objectives within a reasonable time frame is to rely on coordinated multicenter trials.

Multicenter protocols should be conducted in the same manner at all study centers. All procedures should be standardized as much as possible. Investigator meetings should be utilized to reduce site-to-site variation in the treatment regimen or outcome assessment. Personnel should be trained in advance of the trial in key elements of the protocol, including intervention and assessment, with consideration given to periodic re-training during long duration trials (>6 months). Evaluation for any possible nonspecific 'treatment center effect' is aided by recruitment of similar subject numbers at each participating center (ie, 'similar weighting') for analytical purposes. Ongoing monitoring of protocol compliance and the performance of each participating center must be maintained to ensure valid results.

There are additional strengths for a multicenter approach. For instance, subjects are recruited from a wider and more representative population base. Treatments are administered in a wider range of clinical settings (a test of robustness). Thus, multicenter trials can provide a better basis for the subsequent generalization of research findings.

Outcome measures: influence on trial design

Outcome measures have been discussed previously (see SCI Trial Guidelines 2, Steeves *et al.*[1]). However, as a brief review, the primary clinical outcome should be capable of providing the most clinically relevant and convincing evidence directly related to the primary objective of the trial, usually an efficacy measure with distinct thresholds or end points. The protocol should specify the definition of the primary measure, the rationale for its selection, as well as how it will be used in the statistical analysis. The primary outcome measure should have 'face validity': that is, there should be general agreement among knowledgeable experts that the variable is appropriate and adequate to measure the intended outcome, and that it measures a magnitude of change that would be expected to have a meaningful impact on the subject's level of function. Generally, the primary outcome measure should also be utilized to estimate sample size in a power analysis.

Secondary outcome measures include supportive or ancillary measures. These should be limited in number and not be used to estimate the sample size, especially in a Phase 3 trial. Secondary outcome measures in early stage trials can be very useful in defining outcome measures for subsequent components or phases of a study.

Composite scales are those that combine multiple measurements into a single outcome measure based on

a pre-defined algorithm (eg, a rating scale that is the composite of various independent measures). Examples of composite measures that are commonly used in SCI include the Functional Independence Measure (FIM)[18] and the Spinal Cord Independence Measure (SCIM).[19] The creation and definition of composite measures requires that careful consideration should be given to the relative weighting of each measurement within the composite score.

Surrogate measures are indirect measurements of effect in situations where direct measurement of functional clinical benefit is not feasible or practical. For example, anti-hypertensive drugs can be assessed on the basis of reduction in blood pressure, because such a reduction is considered to be a reliable (surrogate) predictor of eventual reduction in clinical functional end points, including heart attack and stroke. In SCI, electrophysiological assessments would be a hypothetical example of a surrogate marker (eg, evoked potentials). Surrogate measures are often used as primary outcome measures in early phase trials to detect the activity of the therapeutics and to allow trials of shorter duration and smaller size to be conducted.[20] Surrogate measures must be reliable predictors of functional benefit, and this represents a challenge in the field of SCI. No commonly accepted surrogate measures exist for SCI clinical trials, but this remains a subject of active clinical research.

Categorized measures consist of analysis of outcome by a dichotomization or other categorization of data, whether continuous or ordinal. In other words, efficacy of a therapeutic 'success' is based on attainment of a prespecified threshold (eg, conversion from an ASIA B classification to ASIA C). This is most useful when the efficacy categories have clear clinical relevance.

In addition to the treatment under investigation, the primary outcome variable may be influenced by other covariates such as neurological level of injury, severity of injury, sex, and age. Treatment covariates such as rehabilitative therapies may also influence the primary outcome variable(s). As an example, a clinical trial of a cellular therapy for SCI, using a functional outcome measure, will need to account for the potential confounding influence of variability of post-treatment rehabilitation. There may also be important outcome differences between subgroups of subjects such as those treated at different centers in a multicenter trial. The design of a clinical trial should consider the range of covariates and subgroups expected to have an important influence on the primary outcome variable, and anticipate how to incorporate these confounding variables into the analysis scheme. When there is a concern regarding the possibility of unequal distribution of important covariates between treatment groups, stratification can be used, allowing for randomization within the defined subgroup.

An acute SCI interventional trial intending to enroll subjects classified as ASIA A, B, C, or D might anticipate a significantly smaller enrolment of motor incomplete (ASIA C and D) subjects than motor

complete (ASIA A and B) subjects. As severity of injury may influence response to treatment, stratification of subjects by motor completeness would be an appropriate method to insure a balanced distribution of this important variable between the treatment groups. The Sygen multicenter trial (Geisler *et al*[14,15]) used this stratification plan as well as stratification by level of injury (cervical or thoracic). The stratification scheme should be defined in the protocol, and the variables used for stratification assignment should be measured before randomization.

Need for control subjects in SCI clinical trials

The use of a control group allows discrimination of patient outcomes owed to the test intervention from outcomes caused by other factors (such as natural progression of SCI, observer or subject bias, or some other covariate). In short, the control group reveals the natural history of patients had they not received the test treatment, or if they received a different treatment known to be effective (an 'active' or 'positive' control). Unfortunately, the lack of adequate controls has been a characteristic of several experimental SCI treatments that have been administered to patient groups in the past. For a more in-depth discussion, see SCI trial Guidelines 3.[3] A brief summary of the various types of control groups is as follows:

Placebo or Sham-Surgery Concurrent Controls are characteristic of a double-blind trial, where an active treatment is compared to an apparently identical treatment that does not contain the test intervention (eg, the NASCIS II trial,[12]). 'Placebo' refers to the administration of an inactive drug, whereas 'sham surgery' refers to a control surgical procedure. In the case of SCI, the nature of a sham surgery procedure could range from a simple skin incision to a more extensive exploration. Placebo or sham surgery groups allow one to keep under control the widest range of non-treatment-related factors that could influence study outcome, including: the natural history of SCI progression, subject and investigator bias, the effect of simply participating in a trial (ie, subjects often show improvement, whether they are in the experimental or control arm of a trial, the placebo effect), influence of another therapy (eg, rehabilitation), and possible subjective elements of diagnosis or assessment.

Best possible treatment (standard of care) Concurrent Control trials consist of assignment of subjects randomly to either an intervention arm or to a no-intervention arm (ie, there is no placebo or sham surgery group), wherein subjects in both arms receive the standard of care treatment. This is not an ideal control because it permits, at best, single-blinded assessment by evaluators who are unaware of a subject's assignment to a treatment or control group. This type of control should only be undertaken when the risk of a sham surgery group is considered unacceptable. There is ongoing debate regarding acceptable risk of sham surgery controls in SCI trials (see SCI Guidelines 3,[3]). Factors

important to this debate include risk to the sham surgery group, benefit to the clarity of outcome in the trial, and societal/SCI-community benefit in gaining knowledge from a well-performed clinical study. At least a sham surgery control must effectively blind the subject, clinical staff, and evaluators, or its purpose will be defeated.

Dose–response Concurrent Control trials consist of subjects who are randomized to one of several fixed-dose treatment groups (where each group is gradually raised to some final fixed dose that is different from the other groups). Between-groups comparisons are made at each group's final fixed dose. This type of trial should be conducted in a double-blind manner. A placebo (vehicle control with a zero dose of the experimental drug) or active control group can also be included in such a design.

Active (Positive) Concurrent Control trials consist of subjects who are randomly assigned to the test treatment or to an active control treatment arm. Once again, this type of trial should be conducted in a double-blind manner. Such trials are often performed as a test of superiority (see above). A critical factor is the capability of this trial design to distinguish an effective benefit from a less effective or no-effect group.

'Baseline' Control trials consist of subjects who serve as their own controls. Their capacity on a relevant outcome measure, during or after the experimental treatment, is compared to the outcome from a previous 'baseline' evaluation. It is dependent on the assumption that the baseline state represents the subject's persistent state in the absence of the test treatment. When the treatment response is dramatic, persistent, and occurs shortly after treatment, it is unlikely to have occurred spontaneously. Nevertheless, such studies are considered as uncontrolled studies and require blinded assessments from independent evaluators. Investigators using 'baseline' controls need to be aware of the limitations of the study, and should be prepared to justify their use.

External Control trials refer to the use of a comparison group that is 'external' to the trial; that is, patients are not prospectively enrolled, treated and assessed within the study protocol. Because external control groups are likely to have significant variance from the study population in important independent variables such as patient characteristics, measurement techniques, and clinical treatments, they are the most problematic controls. There are several types of external control groups: (1) historical controls are patients treated at an earlier time, perhaps from a previous trial or a clinical database or (2) concurrent external controls from a group of patients treated during the same time, but in a different setting or under different circumstances (eg, in different nations). Externally controlled trials tend to overestimate the efficacy of a test therapy. As an accurate comparator group, such controls are very weak due to the different standards of SCI care and rehabilitation that occur between different locations, or they have changed with time.

The reliance on *external controls is a hallmark of an uncontrolled trial*. The estimate of outcomes observed from an external control group should be made conservatively, and only as a guide to possible benchmarks for a larger sample size in a valid, randomized, controlled trial.

Historical SCI control data (ie, untreated patients except for conventional standards of care) can be useful in establishing the natural history of SCI, including the degree of spontaneous recovery and the establishment of initial outcome thresholds that must be achieved to demonstrate the efficacy of an experimental treatment in a randomized, controlled trial. However, historical controls are not useful for prospectively assessing efficacy.

Independent data monitoring committees

For many clinical trials of investigational interventions, especially those that have major public health significance, the responsibility for monitoring comparisons of efficacy and/or safety outcomes should be assigned to an external independent group often called an IDMC or DSMB. The responsibilities of the IDMC should be clearly described in the protocol. The IDMC should be composed of clinicians and scientists knowledgeable in the appropriate clinical trial factors and disciplines, including statistics.

Trial oversight
Ongoing oversight of a clinical trial is a valuable undertaking. It is commonly performed by a Contract Research Organization (CRO) or a Clinical Advisory Board (CAB) that does not require access to information on comparative outcomes or the unblinding of data. Some of the main issues to be tracked are as follows: (1) Is the protocol being followed? (2) Are the collected data of high quality? (3) Are data missing? (4) Are trial design assumptions accurate? (5) Are protocol amendments worthy of consideration? (6) Is subject accrual meeting timeline expectations?

Statistically valid mechanisms for interim analysis, assigning trial participants, and determining sample size power
The number of subjects in a Phase 3 clinical trial should always be large enough to detect a clinically significant difference, if present, between experimental and control groups. The projected enrolment is usually determined by the primary objective of the trial. The method by which the sample size is calculated should be stated in the protocol. The treatment difference to be detected should be based on a judgment concerning the *minimal effect* that has *clinical relevance* in the management of patients. Conventionally, the probability of a type I error (ie, a treatment is concluded to be effective when it actually is not) is set at 5%, and the probability of a type II error (ie, a treatment is concluded to be ineffective when it actually is effective) is set at 10–20%.

'Adaptive' trials are currently being considered as alternative approaches for clinical trial design. Many adaptive trial designs employ Bayesian models, whereby the accumulating results of a trial can be independently assessed at any time, with the possibility of modifying the trial protocol to more efficiently and ethically address the hypothesis under consideration without compromising trial safety. For example, adaptive designs can hypothetically identify, more rapidly, those subjects receiving an ineffective therapeutic dose (including any placebo control), thereby reducing the total number of trial subjects required to reach a statistically valid conclusion. Conversely, the design could also facilitate early and accurate identification of a therapeutic benefit, thereby reducing the risk of withholding a valuable treatment.

This appears to be a potentially valuable approach for SCI clinical trials where the recruitment of a sufficient number of qualified trial participants, within a relatively short time period, can be a challenge. Unfortunately, at present, there is relatively little experience with the use of these statistical approaches, both in terms of design and performance of studies, as well as at the level of regulatory agencies and their familiarity with these approaches. It is to be hoped that the acceptance and appropriate wider use of adaptive methods will be validated in the future.

Data analysis considerations

The principle features of planned statistical analyses of data should be described in the statistical section of any clinical trial protocol. These should include the proposed confirmatory primary outcome analysis as well as the means by which potential problems in data analysis will be managed, such as missing data and protocol violations. The statistical plan should also include a description of the set(s) of subjects that will be used in the main analysis.

The 'Full Analysis Set' is a primary analysis group that includes all randomized subjects, as the term 'intent-to-treat' implies. This is an idealized set, but is often not practically feasible because sizable proportions of subjects often drop out of studies, leading to incomplete data sets. For this reason, exceptions that may occur would justify excluding a randomized subject from the full analysis set. A few examples of such justifiable exclusions would be the discovery of eligibility violations that were objectively measured before randomization, failure to receive any of the study treatments, or a complete lack of post-randomization data.

The 'Per Protocol' set is a subgroup of the Full Analysis subjects who meet more of the protocol definitions for compliance with the treatment regimen, control of defined covariates (such as rehabilitation therapies or concomitant medications), and data collection. Criteria for inclusion in the Per Protocol set might include delivery of a predetermined proportion of the treatment or completion of a predetermined portion of the primary outcome variable measurements. The Per

Protocol Set is sometimes referred to as the set of 'valid cases', the 'efficacy sample', or the 'evaluable subjects'.

In confirmatory Phase 3 trials, it is usually appropriate to plan to conduct both an analysis of the Full Analysis set and a Per Protocol analysis, so that any differences between these analyses can be the subject of explicit discussion and interpretation. When the Full Analysis Set and Per Protocol Set lead to essentially the same conclusions, confidence in the trial result is increased.

Missing values are an important potential source of bias in clinical trials. All trials will have missing data, and it is very important to indicate *a priori* how missing data will be managed in a trial. A trial may be regarded as valid, provided the methods of dealing with missing values are sensible and predefined in the protocol. Outliers are data points so far removed from other values that their presence cannot be attributed to a simple chance occurrence. These values are usually at least 2 SD from the mean. Outliers should be strictly defined *a priori*, and the basis for their definition must be justified on clinical and statistical grounds. Further, means of managing outlying data must not favor any treatment group *a priori*.

Registration of clinical trials

Prompted by concerns over selective reporting of clinical trial results, the International Committee of Medical Journal Editors (ICMJE) has recently promoted the *a priori* registration of clinical trials with a recognized public registry as a requirement for publication.[21] In concert with the World Health Organization, the ICMJE has developed a minimal registration dataset that is recommended for adoption by trial registries. This data set includes trial name, sponsorship, ethics board/Institutional Review Board approval, inclusion/exclusion criteria, target sample size, primary outcome, and key secondary outcomes.

The registration of a trial will require that the sponsor and investigators 'go on record' to declare the basic elements of the research protocol before initiation of enrolment. this process will not only enable public awareness of the existence of clinical trials, but also it will discourage *post hoc* redefinition of key research design elements such as primary outcome, target enrolment, and eligibility criteria. For these reasons, registration of interventional trials in SCI should be strongly encouraged, especially in the case of Phase 2 and Phase 3 studies. A widely used clinical trial registration system can be found at www.clinicaltrials.gov.

Summary

- SCI Trials should be designed in accordance with the ICH Clinical Trial guidelines.
- SCI Trials should obtain the appropriate regulatory approval.

- SCI Trials should utilize well-defined protocols that include prospectively defined and appropriate Inclusion/Exclusion criteria, a complete description of interventions and procedures, and a detailed statistical analysis plan.
- SCI Trials should include ongoing data quality monitoring and the use of an IDMC.
- SCI Trials should register the protocol to enable subsequent publication of trial results in peer-reviewed journals.
- SCI Trial designs will need to address the problem that small numbers of subjects are available for study at acute and subacute stages of injury (ie, low incidence). Means of managing this issue include:

  ○ Multicenter trial designs.
  ○ RCT designs with placebo or sham surgery control groups whenever feasible, not just when convenient.

- SCI Trial designs should take into account the fact that a substantial proportion of subjects enrolled in the acute (and even subacute) time frame will experience spontaneous improvement (ie, the natural history of recovery), depending on the initial extent of injury.
- The use of external controls in SCI clinical trials is strongly discouraged.
- SCI Trials of therapies intended to improve neurological outcomes should measure not only the *magnitude* of change in the primary endpoint but also the *duration* of improvement.

  ○ Claims of lasting therapeutic effect on neurological function should be supported by follow-up data extending to at least 1 year.

- SCI Trial design should carefully consider the selection of primary outcome measures:

  ○ The adequacy of current measures should be carefully assessed, together with the need for development of novel validated tools.
  ○ Surrogate or composite variables as *primary* outcome measures in SCI trials are either nonexistent or have yet to be adequately validated.

- SCI trial designs should consider the control of a multiplicity of confounding independent variables that are:

  ○ Center and treatment related (eg, in multicenter trials),
  ○ Patient related: pre-morbid and injury related (level and severity, etc).
  ○ Related to interactions in combination therapies

- To manage confounding variables, SCI clinical trials should:

  ○ Whenever possible in the evaluation of efficacy, use randomized, controlled trial designs. Noncontrolled or externally controlled trials will, in general, be appropriate only for early phases of treatment development.

  ○ When placebo or sham surgery controls and double blinding cannot be included, at least single-blinded outcome assessments should be incorporated into trial design.
  ○ Consider factorial design.
  ○ Consider add-on design.
  ○ Carefully consider sample size implications.
  ○ Carefully note the potential for placebo effects.

The field of SCI interventional research for improved neurological recovery has witnessed significant advances over the past several decades. Trials of neuroprotective pharmacological strategies provide examples of the evolution of clinical trial design. The advent of cellular therapies and other novel interventions, holds much promise, but will pose additional challenges for clinical trial design. The ICCP clinical trials panel proposes the set of guidelines published in this issue of Spinal Cord in an effort to promote the best possible clinical science with the goal of facilitating the development of truly effective and safe treatments for people with SCI.

Glossary of definitions

(Additional glossaries are included in the three accompanying papers).

DSMB is the abbreviation for Data Safety Monitoring Board.

IDMC stands for Independent Data Monitoring Committee.

ICH is the International Conference on harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) brings together the regulatory authorities of Europe, Japan, and North America with experts from the pharmaceutical industry to discuss scientific and technical aspects of product registration. The purpose is to make recommendations on ways to achieve greater harmonization in the interpretation and application of technical guidelines and requirements for product registration. The objective of such harmonization is a more economical use of human, animal, and material resources, and the elimination of unnecessary delay in the global development and availability of new medicines while maintaining safeguards on quality, safety and efficacy, and regulatory obligations to protect public health http://www.ich.org/.

Helsinki Declaration or the Declaration of Helsinki, developed by the World Medical Assembly, is a set of ethical principles for the medical community regarding human experimentation. It was originally adopted in June 1964 and has since been amended many times. The recommendations concerning the guidance of physicians involved in medical research may be found on http://www.wma.net/e/policy/b3.htm.

Belmont Report is a report created by the former United States Department of Health, Education, and Welfare (which was renamed to Health and Human Services) entitled 'Ethical Principles and Guidelines for the Protection of Human Subjects of Research. The text

is available on http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm.

RCT stands for Randomized Clinical Trial.

Translational Research is the necessary pre-clinical research specifically designed to answer important questions of dosing, delivery methodology, timing, and functional outcome of a therapeutic in anticipation of and before human trials.

Type 1 error is the chance that the study hypothesis is falsely accepted.

Type 2 error is the chance that the study hypothesis is falsely rejected.

References

1 Steeves JD et al. Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP Panel: Clinical trial outcome measures. Spinal Cord 2007; **45:** 206–221 (this issue).

2 Fawcett JW et al. Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP Panel: Spontaneous recovery after spinal cord injury and statistical power needed for therapeutic clinical trials. Spinal Cord 2007; **45:** 190–205 (this issue).

3 Tuszynski MH et al. Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP Panel: Clinical trial inclusion/exclusion criteria and ethics. Spinal Cord 2007; **45:** 222–231 (this issue).

4 Clifton GL et al. Omental transposition in chronic spinal cord injury. Spinal Cord 1996; **34:** 193–203.

5 Halter JA, Blight AR, Donovan WH, Calvillo O. Intrathecal administration of 4-aminopyridine in chronic spinal injured patients. Spinal Cord 2000; **38:** 728–732.

6 Segal JL et al. Absorption characteristics of sustained-release 4-aminopyridine (fampridine SR) in patients with chronic spinal cord injury. J Clin Pharmacol 2000; **40:** 402–409.

7 Knoller N et al. Clinical experience using incubated macrophages as a treatment for complete spinal cord injury: Phase I study results. J Neurosurg Spine 2005; **3:** 173–181.

8 Shapiro S et al. Oscillating field stimulation for complete spinal cord injury in humans: a Phase 1 trial. J Neurosurg Spine 2005; **2:** 3–10.

9 Freed CR et al. Transplantation of embryonic dopamine neurons for severe Parkinson's disease. N Engl J Med 2001; **344:** 710–719.

10 Olanow CW et al. A double-blind controlled trial of bilateral fetal nigral transplantation in Parkinson's disease. Ann Neurol 2003; **54:** 403–414.

11 Bracken MB et al. Efficacy of methylprednisolone in acute spinal cord injury. JAMA 1984; **251:** 45–52.

12 Bracken MB et al. A randomized controlled trial of methylprednisolone or naloxone in the treatment of acute spinal cord injury: results of the second national acute spinal cord injury study. N Engl J Med 1990; **322:** 1405–1411.

13 Bracken MB et al. Administration of methylprednisolone for 24 or 48 hours or tirilazad mesylate for 48 h in the treatment of acute spinal cord injury: results of the third national acute spinal injury randomized controlled trial. JAMA 1997; **277:** 1597–1604.

14 Geisler FH, Dorsey FC, Coleman WP. Recovery of motor function after spinal cord injury: a randomized, placebo-controlled trial with GM-1 ganglioside. N Engl J Med 1991; **324:** 1829–1838.

15 Geisler FH, Coleman WP, Grieco G, Poonian D. The Sygen multicenter acute spinal cord injury study. Spine 2001; **26:** S87–S98.

16 Potter PJ et al. Randomized double-blind crossover trial of Fampridine-SR (sustained release 4-aminopyridine) in patients with incomplete spinal cord injury. J Neurotrauma 1998; **10:** 837–849.

17 Cardenas DD et al. Phase 2 trial of sustained-release Fampridine in chronic spinal cord injury. Spinal Cord [E-pub ahead of print: 13 June 2006; doi:10.1038/sj.sc.3101947].

18 Hamilton B, Granger C, Sherwin F. A uniform national data system for medical rehabilitation. In: Fuhrer M (ed). Rehabilitation Outcomes: Analysis and Measurement. Baltimore: Brooks 1987, pp 137–147.

19 Catz A, Itzkovich M, Agranov E, Ring H, Tamir A. SCIM – spinal cord injury independence measure: a new disability scale for patients with spinal cord lesions. Spinal Cord 1997; **35:** 850–856.

20 Katz R. Biomarkers and surrogate markers: an FDA perspective. NeuroRx 2004; **1:** 189–195.

21 DeAngelis C et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. Ann Intern Med 2004; **141:** 477–478.