Published in final edited form as:

J Biopharm Stat. 2009 ; 19(1): 204–216. doi:10.1080/10543400802536271.

A Decision Algorithm for Translating Preclinical Trial Results to Enhance Recovery after Stroke

Jonathan D. Mahnken^{a,b,c}, Matthew S. Mayo^{a,b}, and Randolph J. Nudo^{c,d}

a Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, USA

b Center for Biostatistics and Advanced Informatics, University of Kansas Medical Center, Kansas City, KS, USA

c Landon Center on Aging, University of Kansas Medical Center, Kansas City, KS, USA

d Department of Molecular and Integrative Physiology, University of Kansas Medical Center, Kansas City, KS, USA

Abstract

A decision algorithm was required to evaluate the first half of a cooperative agreement for preclinical trials to optimize medical device parameters to enhance stroke recovery. Continued funding was contingent upon the midpoint evaluation, called the milestone decision. We developed an algorithm, which summarized our rodent and primate model results. Primary outcomes weighed more heavily than secondary outcomes, and the primate model more heavily than rodent models. By controlling the type I error for this milestone decision, the probability of correctly discontinuing the study if treatment was not beneficial was high (>0.84). Similar algorithms may be adapted for other milestone-driven projects.

Keywords

cooperative agreement; milestones; translational research; type I error

INTRODUCTION

Each year in the U.S., 700,000 people experience a stroke. While the majority of individuals survive, about 15–30% of stroke survivors are permanently disabled, making stroke the leading cause of motor dysfunction in the adult population (American Heart Association, 2006). Of stroke survivors over age 65, 50% experience hemiparesis, 30% are unable to walk without assistance, and 26% are institutionalized in nursing homes (Kelly-Hayes, Beiser, Kase, 2003). As of 2003, the prevalence of stroke survivors in the U.S. was 5.5 million (American Heart Association, 2006).

Several potential treatment approaches now hold promise for restoring motor function even in chronic stages after stroke (Wolf, Thompson, Morris, 2005; Goldstein, 2003). One novel intervention that has shown efficacy in animal stroke models and in pilot clinical trials is the use of direct electrical stimulation of the cerebral cortex combined with physical rehabilitation (Brown, Lutsep, Weinand, Cramer, 2006; Adkins, Campos, Quach et al., 2006; Plautz, Barbay, Frost, et al., 2003; Teskey, Flynn, Goertzen, Monfils, Young, 2003; Kleim, Bruneau,

VandenBerg, et al., 2003). However, the effects of electrical stimulation on cortical plasticity and post-injury recovery have rarely been examined systematically. To more fully understand the potential for this novel mode of treatment, investigators from the University of Kansas Medical Center, the University of Texas-Austin, the University of Calgary, and the University of Florida are collaborating on research designed to optimize the stimulation parameters to enhance functional recovery through a series of preclinical trials performed in rodent and nonhuman primate models. This research is supported by a cooperative agreement from the National Institute of Neurological Disorders and Stroke (NINDS) ("Cortical Stimulation to Enhance Recovery after Stroke," PI: Nudo RJ, 1-U54-NS-48126-1A1). The funding mechanism for this research has in place a milestone decision halfway through the study (i.e., at the end of year two). This milestone is in place to ensure that progress towards optimizing cortical stimulation device protocols for stroke patients warrants continued support from this cooperative agreement for the subsequent proposed preclinical trials. Such milestone-driven cooperative agreements are becoming more common under recent initiatives contained in the National Institutes of Health Roadmap for Medical Research (National Institutes of Health, 2006).

The basic hypothesis is that changes in motor cortical excitability induced by cortical stimulation will provide a functional advantage for the recovering stroke survivor, presumably by increasing the excitability of cortical neurons during the time of rehabilitative therapy. Varying the treatment protocol for both the cortical stimulation as well as the corresponding rehabilitative training in the animal models will provide insight into how these devices may be applied to human subjects in clinical trials. Because of the numerous parameters that may be varied using electrical stimulation (e.g., frequency, pulse width, pulse pattern, etc.), it is more reasonable to proceed through a series of optimization studies in animal models to narrow the range of potential options before proceeding to human trials. The basic framework is to identify optimal parameters in the rodent models, to translate these parameters into the nonhuman primate model, and, in later clinical trials, to translate these findings into human application. Translation from rodent to primate will provide insight—albeit preclinical—into the safety and efficacy of this innovative treatment device in humans. The framework for this program is as follows.

Milestone 0: Optimize primary cortical stimulation parameters in rodents

Milestone 1: Generalize primary cortical stimulation parameters to nonhuman primates; optimize secondary cortical stimulation parameters and primary rehabilitative therapy parameters in rodents

Milestone 2: Generalize secondary cortical stimulation parameters and primary rehabilitative therapy parameters to primates; optimize tertiary cortical stimulation parameters and secondary rehabilitative therapy parameters in rodents

Optimized primary cortical stimulation parameters were derived in rodent models (Milestone 0) in earlier studies funded by Northstar Neuroscience, Inc. (Seattle, Washington) to provide preliminary support for this program. Upon completion of the first two years, evaluation of the translation of these parameters to nonhuman primates and of new parameters tested in rodents was conducted (Milestone 1). Further translation to primates of Milestone 1 findings and new parameter optimization in rodent models were proposed for years 3–4. Results of these experiments will be evaluated at Milestone 2. Thus, Milestone 1 was the critical milestone for the decision to extend funding to years 3–4. These milestone evaluations served not only to guide the direction of the subsequent preclinical trials, but also provide insight into the optimal application of this device for human clinical trials.

METHODS

As part of the program, which consisted of three separate projects, specific rules by which the milestone decisions were to be based were requested by the NINDS. The guidelines for proceeding to years 3–4 preclinical studies needed to be explicitly stated prior to funding. This decision was to be based on the culmination of evidence from the two rodent models (Projects 1 and 2) and the nonhuman primate model (Project 3), taking into account not only statistical significance but also non-significant trends showing improved performance among treatment groups. It was extremely important to not overlook trends lacking statistical significance at the conventional level (i.e., those with $P > 0.05$) that may be indicative of the potential for improved functional status among human stroke survivors. This was particularly important in the nonhuman primate projects that had a smaller sample. We chose to balance this tradeoff between type I and type II errors by incorporating test results with higher type I errors than the typical 0.05 threshold, thereby increasing the statistical power of these tests (i.e., reducing the type II error rates). To quantify the process, endpoint decision rules were incorporated to include a broader range of outcomes and clinical relevance into these important decisions. These decision rules treated primary outcomes as stronger evidence of a treatment effect than secondary outcomes and the primate model as stronger evidence than rodent models.

For each of the projects the primary and secondary outcomes were defined (Table 1). Treatment efficacy was determined by comparing each treatment group to the control for the primary and secondary outcomes. To determine persistence, a single, primary endpoint was evaluated by, again, comparing treatment groups to the control. Experimental treatment regimens that showed efficacy and persistence over controls supported the decision to carry on with the planned research (years 3–4). The translation supported by this program began in rodents and moved on to nonhuman primates. The process of how the primary outcomes versus secondary outcomes derived the efficacy and persistence interpretations within a given model (rodent or primate) and how the efficacious/persistent rodent versus primate therapies were evaluated are presented below.

Analysis of Rodent Model from Project 1

The rodent model from Project 1 compared multiple treatment levels against the control group for each outcome. The primary outcome and reaching success on any other day of testing during treatment secondary outcomes (Table 1) were analyzed using analysis of variance (ANOVA) to test for an overall treatment effect. This was followed by Dunnett's test (Kuehl, 1994)—contingent on the results of the overall ANOVA test—to compare each treatment level to the controls. The slope of recovery curve (secondary) outcome (Table 1) was tested using linear contrasts of parameter estimates from a random-coefficients regression model (Raudenbush, Bryk, 2002) with a random intercept to account for the variation in skill levels of each rodent and for the correlation between measures within animals. The contrasts compared the expected changes from baseline to final follow-up in the treated versus control groups.

Analysis of Rodent Model from Project 2

The rodent model from Project 2 was similar to that of Project 1, and compared multiple treatment levels against the control group for each endpoint. The primary outcomes (efficacy and persistence) and secondary efficacy outcomes reaching success on any other day of testing during treatment secondary endpoints (Table 1) were analyzed using ANOVA to test for an overall treatment effect and Dunnett's test for multiple comparisons. The slope of recovery curve (secondary efficacy) outcome (Table 1) was tested using parameter estimates from a random-coefficients regression model as described above.

Analysis of Primate Model from Project 3

The primate model from Project 3 compared the treatment group to the control group for each endpoint. The primary outcomes (efficacy and persistence) and secondary (efficacy) outcomes average performance time on any other day of testing during treatment and reaching success on any other day of testing during treatment (Table 1) were analyzed using the Wilcoxon rank sum test (Hollander, Wolfe, 1999) due to the small sample size. The slope of recovery curve (secondary efficacy) outcome (Table 1) was tested using parameter estimates from a random-coefficients regression model. The days to criterion performance (secondary efficacy) outcome was compared between the treatment and control groups using the log-rank test (Klein, Moeschberger, 1997) to accommodate the right-censoring in these data.

Efficacy and Persistence Algorithms

Table 2 represents the algorithm describing how efficacy of a treatment was quantitatively determined. If the primary outcome was significant ($P \leq 0.05$), the treatment was considered efficacious. When the primary outcome was near significance ($0.05 < P \leq 0.2$), the treatment was considered efficacious if at least one secondary outcome was also either near significance or significant ($P \leq 0.2$). Finally, if there were at least two significant secondary outcomes ($P \leq 0.05$), then the treatment was deemed efficacious. These P-value thresholds were based on two-sided tests. This algorithm allowed for consideration of trends as opposed to relying entirely on a single statistical test, while retaining a quantitative decision process for determining efficacy. The secondary outcomes listed below incorporated not only measures at specific points in time, but also the rate of recovery by using regression models that accounted for correlation between measures within animals. The secondary outcomes should and did play a role in determining the efficacy of a treatment. As each model had at most one persistence endpoint, a treatment was deemed persistent if it had $P \leq 0.05$. The pre-determined follow-up time used for testing persistence varied across the projects in this program depending on both the aims and the type of model (rodent or primate).

Two-sided Versus One-sided Tests

The thresholds selected for significance (0.05) and near significance (0.20) were set based on two-sided tests. Under the null hypothesis of no effect versus the alternative of any effect, half of the two-sided test rejected would be rejected when the treatment group outperformed the control group and the other half when the control group outperformed the treatment group. However, decisions in support of proceeding with years 3–4 of the program from our algorithm (Table 2) were based on those meeting the significance/near significance threshold(s) and having the treatment group demonstrate a benefit over the controls, akin to a clinical trial. Thus, the resulting type I error rates for these significance/near significance decisions were actually half that of the values presented above (i.e., 0.025 and 0.10 for significance and near significance, respectively).

Selection of Cut-points

Various choices of the cut-points for P-values in the algorithm described above were given consideration. For the lower P-value indicative of statistical significance the traditional $\alpha = 0.05$ level was chosen. This helped reduce the possibility of erroneously concluding a parameter was efficacious and/or persistent when, in fact, it is not. However, clinically relevant trends can be overlooked with such low α -levels (Rothman, Greenland, 1998). The “near significance” cut-point was set higher to $\alpha = 0.2$ to protect against excluding factors that are truly relevant in these small studies. Such findings are treated as weaker evidence of a treatment effect in the milestone decision process, though, relative to those that are statistically significant.

Milestone Decision to Continue Program

Summarizing years 1–2 of the program as a whole incorporated scientific relevance in the decision-making process, as implications from rodent models were different from those of nonhuman primate models in terms of translation into clinical significance. At the end of year two, if parameters had demonstrated efficacy or persistence in the nonhuman primate model or in both rodent models, the decision was to continue as proposed in the program application. While heavier emphasis was placed on the nonhuman primate model, promising trends observed only in the rodent models would not be overlooked based strictly on the nonhuman primate model. This quantitative approach helped guide the necessary decision to proceed or discontinue the study at the program milestone.

Type I Error of Milestone Decision to Continue Program

Defining efficacy and persistence events for the i th project (Table 1) as E_i and P_i , the probability of proceeding to years 3–4 of this study was

$$\begin{aligned} & \Pr\{[E_1 \cap (E_2 \cup P_2)] \cup (E_3 \cup P_3)\} \\ = & \Pr(E_1) \cdot \Pr(E_2 \cup P_2) + \Pr(E_3 \cup P_3) - \Pr(E_1) \cdot \Pr(E_2 \cup P_2) \cdot \Pr(E_3 \cup P_3). \end{aligned} \quad (1)$$

To determine the upper bound of the milestone decision type I error rate, we calculated these probabilities under the null hypothesis of no benefit due to cortical stimulation therapy and assumed all tests of efficacy and persistence were pairwise independent.

As noted above, the P -value thresholds defined in Table 2 were based on two-sided tests, but only those in the direction of a benefit due to cortical stimulation were used for evidence of proceeding with years 3–4 of this cooperative agreement. Thus, the type I error of each individual test was actually half that presented in Table 2.

Starting with the persistence outcome from project 3, $\Pr(P_3) = 0.025$. As there were no secondary persistence measures, persistence relied entirely on the 12 week performance time measure (Table 1). The probability of a type I error with respect to E_3 was more complex. The probability that a P -value for the primary endpoint was ≤ 0.05 in the correct direction was 0.025. Under the null hypothesis, this primary endpoint would have $0.05 < P \leq 0.2$ in the proper direction with probability 0.075. This value was multiplied by the probability that at least one secondary endpoint would have $P \leq 0.2$. Now, the probability that at least one secondary endpoint would have $P \leq 0.2$ was bounded above by $1 - (1 - 0.1)^{32} \approx 0.9657$. This value was determined by finding $\Pr(Y > 0)$ where $Y \sim \text{Binomial}(32, 0.1)$ and 0.1 represented the probability that any one of the 32 tests (29 tests of average performance time during the treatment period, one for slope of the recovery curve, one for reaching success on the last day of treatment, and one for days to criterion performance) would have $P \leq 0.2$ in the correct direction under the null hypothesis. As indicated, this probability was actually the upper bound because, in the event that these measures were not independent, the probability that at least one secondary endpoint would have $P \leq 0.2$ in the correct direction would be < 0.9657 . We based this argument heuristically on the fact that, if these events were perfectly correlated, then they would essentially reduce to a single event, and the probability of such an event having $P \leq 0.2$ in the correct direction under the null hypothesis would have been 0.1; so, our calculations represented a conservative estimate of this contribution to the type I error of the decision algorithm. Thus, the probability of finding cortical stimulation efficacious in the primate model when $0.05 < P \leq 0.2$ for the primary endpoint and $P \leq 0.2$ for at least one secondary endpoint was approximately $(0.075)(0.9657) \approx 0.0724$. Finally, when the primary efficacy had $P > 0.2$ (or $P \leq 0.2$ but in the wrong direction for efficacy) cortical stimulation will be deemed efficacious if at least two secondary endpoints had $P \leq 0.05$. The upper bound for this

probability (that two or more secondary endpoints had $P \leq 0.05$) was found using probability generating functions and convolution (Chiang, 1980). In brief, we let Y_1 be the event that average performance time on at least one day of testing (other than the last day of testing) during treatment would have $P \leq 0.05$ in the correct direction (Table 1). By similar arguments as above, $\Pr(Y_1 > 0)$ was bounded above by $1 - (1 - 0.1)^{29} \approx 0.5201$. We let Y_2 be the total number of the other three secondary endpoints (Table 1) with $P \leq 0.05$ in the correct direction. Conservatively assuming independence, using the same heuristic argument as above, $Y_1 \sim \text{Bernoulli}(0.5201)$, $Y_2 \sim \text{Binomial}(3, 0.025)$, and Y_1, Y_2 were assumed independent. So then, by convolution the probability generating function for $Y_1 + Y_2$ indicated that

$$\Pr(Y_1 + Y_2 \geq 2) \approx 1 - (1 - 0.5201)(1 - 0.025)^3 - [(0.5201)(1 - 0.025)^3 + 3(1 - 0.5201)(0.025)(1 - 0.025)^2],$$

i.e., $\Pr(Y_1 + Y_2 \geq 2) \approx 0.0389$. So then, the probability of concluding efficacy when the primary efficacy endpoint had $P > 0.2$ (or had $P \leq 0.2$ but in the wrong direction) would occur with probability $(0.9)(0.0389) \approx 0.035$ under the null hypothesis. Taken collectively,

$$\Pr(E_3) \approx 0.025 + 0.0724 + 0.035 = 0.1324$$

and

$$\Pr(E_3 \cup P_3) = \Pr(E_3) + \Pr(P_3) - \Pr(E_3) \cdot \Pr(P_3) \approx 0.1541.$$

We determined the probabilities for project 2 under the null hypothesis in a similar fashion by again assuming the tests were pairwise independent. As with project 3, there was only a primary endpoint for persistence, thus, $\Pr(P_2) = 0.025$. The probability of determining efficacy in this model based solely on the results of the test for the primary efficacy endpoint was 0.025 under the null hypothesis. The probability that the primary endpoint would have $0.05 < P \leq 0.2$ and at least one secondary endpoint would have $P \leq 0.2$ was bounded above by $(0.075)\{[1 - (1 - 0.1)^5]\} \approx 0.0307$ based on similar arguments as described for project 3 (the reaching success on any other day of testing was assessed on four days). Claiming efficacy for the rodent model of project 2 on the basis of at least two secondary endpoints (i.e., all secondary endpoint in this model) having $P \leq 0.05$ (when $P > 0.2$ for the primary endpoint or $P \leq 0.2$ in the wrong direction) would occur with approximate probability $(0.9)(0.0963)(0.025) \approx 0.0022$, with the latter two values on the left hand side being the probabilities of $P \leq 0.05$ for the reaching success on any other day of testing during treatment and slope of recovery curve secondary endpoints, respectively (Table 1). Taken collectively,

$$\Pr(E_2) \approx 0.025 + 0.0307 + 0.0022 = 0.0579$$

and

$$\Pr(E_2 \cup P_2) = \Pr(E_2) + \Pr(P_2) - \Pr(E_2) \cdot \Pr(P_2) \approx 0.0814.$$

Project 1 did not have any persistence endpoints. The probability of determining efficacy in this model based on the primary efficacy endpoint was 0.025. The probability that the primary endpoint would have $0.05 < P \leq 0.2$ and at least one secondary endpoint with $P \leq 0.2$ was bounded above by $(0.075)\{[1 - (1 - 0.1)^{10}]\} \approx 0.0488$ based on similar arguments to those for project 3 (the reaching success on any other day of testing was assessed on nine days). Claiming efficacy for the rodent model of project 1 on the basis of both secondary endpoints having $P \leq 0.05$ (with $P > 0.2$ for the primary endpoint or $P \leq 0.2$ but in the wrong direction) would occur with approximate probability $(0.9)(0.2038)(0.025) \approx 0.0046$. Taken collectively,

$$\Pr(E_1) \approx 0.025 + 0.0488 + 0.0046 = 0.0784.$$

Using these values, the assumption of pairwise independent tests, and equations (1) and (2), the type I error for this decision to proceed was calculated.

$$\Pr[(E_1 \cap (E_2 \cup P_2)) \cup (E_3 \cup P_3)] \approx 0.1596$$

In words, the probability of proceeding with years 3–4 of this study when there was no benefit from cortical stimulation was 0.1596. Put differently, if there was, in fact, no benefit due to cortical stimulation, there was over an 84% chance that the algorithm would have indicated that the study should be discontinued at the end of year 2.

In addition to this upper bound for the milestone decision type I error, a lower bound was also determined theoretically. This result assumed perfect correlation between tests for primary and secondary outcomes within projects. That is, within a given project, if any one of the primary (efficacy or persistence) or secondary outcomes was significant, then all of the other outcomes within that project were significant. Similarly, if one was near significance, all outcomes within that project were near significance. Independence between tests from different project was retained in this calculation, as tests of outcomes from different projects were based on different animals from different laboratories. Heuristically, the lower bound for this type I error was found by treating each project as though it had a single response. Beginning with Project 3, all outcome measures would have tests with $P \leq 0.05$ and in the correct direction with probability 0.025; and all tests would have $0.05 < P \leq 0.2$ and in the correct direction with probability 0.075. These represent the scenarios that indicate proceeding with years 3–4 when the null hypothesis is true for perfectly correlated tests within projects. Thus, the $\Pr(E_1)$, $\Pr(E_2 \cup P_2)$, and $\Pr(E_3 \cup P_3)$ are all equal to 0.1. Using this result and equation (1), the lower bound for the type I error can be calculated as

$$\Pr(E_1) \cdot \Pr(E_2 \cup P_2) + \Pr(E_3 \cup P_3) - \Pr(E_1) \cdot \Pr(E_2 \cup P_2) \cdot \Pr(E_3 \cup P_3) = 0.1090.$$

Thus, the type I error for this algorithm is bounded by the interval [0.1090, 0.1596].

Simulation Study

A simulation study was conducted to evaluate this algorithm under various cut-points for the significance and near-significance levels. These simulations allowed for correlation between tests of outcomes (primary and secondary; efficacy and persistence) within projects. This was accomplished by generating multivariate binary data that allowed for correlation between the binary variates. We used the method of Qaqish (2003) to create these data with an exchangeable correlation matrix, which allowed for simulating correlated tests within projects. Specifically,

for each simulated set of test results a (pseudo-)random vector of uniform(0, 1) variates was generated for each outcome over all of the projects in the program. These vectors were used to determine whether the each simulated outcome resulted in a significant test or not based on equation (5) of Qaqish (2003), which allowed for an exchangeable correlation between tests within each project. In other words, these vectors of uniform(0, 1) variates were used to create a new set of vectors that contained the correlated binary data. Using this same simulated (pseudo-)random multivariate uniform(0, 1) vector, additional correlated binary vectors were derived to simulate test results at the near-significance level—again using equation (5) of Qaqish (2003). The results of these simulated correlated tests were then used to determine whether the algorithm would indicate proceeding with years 3–4 or not. These steps were repeated 10,000 times for each combination of near-significance (0.10, 0.15, 0.20, and 0.25) and level of exchangeable correlation between tests within a project (0.00, 0.25, 0.50, 0.75, and 1.00) parameters. Mathematica 6.0.1.0 (Wolfram Research, Inc., 1998–2007) was used to perform this simulation study.

RESULTS

Results of the simulation study are presented in Table 3. As described above, the boundaries for the milestone decision type I error were [0.1090, 0.1596], with the lower bound representing the case of perfect correlation between test results within projects and the upper bound representing the case of complete independence between test results. These compare to values of 0.1095 and 0.1581 for the simulation results using these same significance (0.05) and near-significance (0.20) parameters under perfect correlation ($\rho = 1.00$) and independence ($\rho = 0.00$), respectively. These simulation results demonstrate a decrease in the algorithm type I error with increased correlation between tests within projects for given levels of significance and near-significance.

DISCUSSION

We developed an algorithm to be used to quantify the results of related preclinical trial findings from a National Institutes of Health- (NIH-)funded translational research program. This algorithm was laid out as our objective and operationally-defined criteria for success in terms of the first two years of this four-year cooperative agreement. The type I error in the decision to proceed with years 3–4 of this study was bounded by [0.1090, 0.1596], so there was an 84–89% chance that the program would cease prior to years 3–4 if the cortical stimulation treatment in these preclinical studies was not effective.

The algorithm presented here is unique in two ways. First, it combined data from two different animal species to inform investigators whether a particular intervention should be considered for human clinical trials, and what parameters should be used. Second, it was designed specifically for NIH program officials to use in evaluating whether the project should continue beyond a predetermined milestone. However, other milestone-driven grants are becoming more common as NIH concentrates more heavily on translational research programs. Similar algorithms may easily be adapted using the model presented here. Modifications to cut-point levels, the number of endpoints, and the number of models can be incorporated to derive midpoint guides specific to other projects, and the type I error of these decision algorithms can be assessed as we have presented here. Such algorithms can provide formal, objective criteria by which operating characteristics of these important decisions, such as type I error, can be identified and controlled.

The purpose of this research is to better understand how to affect post-stroke recovery through systematic preclinical trials examining various parameters of cortical electrical stimulation combined with rehabilitative training. Such translational research initiatives from basic science

to clinical practice are a part of the NIH Roadmap for re-engineering the clinical research enterprise (National Institutes of Health, 2006). Through this type of collaboration, expertise can be garnered under a common research theme to enhance therapy and improve motor-function outcomes among the many survivors of stroke, as well as reduce the lag time between the breakthroughs in basic science and the implementation of subsequent, innovative treatments for disease.

Acknowledgments

Grant Support Acknowledgement: This research was supported by a cooperative agreement from the National Institute of Neurological Disorders and Stroke, National Institutes of Health (1-U54-NS-48126-1A1).

The authors thank Drs. Theresa Jones, Jeffrey Kleim, Erik Plautz, and G. Campbell Teskey for their helpful comments in creating the algorithm. The authors also thank the reviewers for their insight, which has strengthened this work.

References

- Adkins DL, Campos P, Quach D, et al. Epidural cortical stimulation enhances motor function after sensorimotor cortical infarcts in rats. *Exp Neurol*. 2006Epub ahead of print
- American Heart Association. Heart Disease and Stroke Statistics — 2006 Update. Dallas, Texas: American Heart Association; 2006.
- Brown JA, Lutsep HL, Weinand M, Cramer SC. Motor cortex stimulation for the enhancement of recovery from stroke: a prospective, multicenter safety study. *Neurosurgery* 2006;58(3):464–473. [PubMed: 16528186]
- Chiang, CL. An Introduction to Stochastic Processes and Their Applications. Huntington, NY: Robert E. Krieger Publishing Company, Inc; 1980. p. 31-34.
- Goldstein LB. Pharmacotherapy in Stroke Rehabilitation. *Adv Neurol* 2003;92:447–450. [PubMed: 12760212]
- Hollandar, M.; Wolfe, DA. Nonparametric Statistical Methods. Vol. 2. New York, NY: John Wiley & Sons, Inc; 1999. p. 106-125.
- Kelly-Hayes M, Beiser A, Kase CS, et al. The influence of gender and age on disability following ischemic stroke: the Framingham study. *J Stroke Cerebrovasc Dis* 2003;12(3):119–26. [PubMed: 17903915]
- Kleim JA, Bruneau R, VandenBerg P, et al. Motor cortex stimulation enhances motor recovery and reduces peri-infarct dysfunction following ischemic insult. *Neurol Res* 2003;25(8):789–793. [PubMed: 14669520]
- Klein, JP.; Moeschberger, ML. Survival Analysis: Techniques for Censored and Truncated Data. New York, NY: Springer-Verlag; 1997. p. 191-202.
- Kuehl, RO. Statistical Principles of Research Design and Analysis. Belmont, CA: Duxbury Press; 1994. p. 93-95.
- National Institutes of Health. Re-engineering the Clinical Research Enterprise. National Institutes of Health; Bethesda, MD: [accessed 3 February 2006]. Available from <http://nihroadmap.nih.gov/clinicalresearch/overview-translational.asp>
- Plautz EJ, Barbay S, Frost SB, et al. Post-infarct cortical plasticity and behavioral recovery using concurrent cortical stimulation and rehabilitative training: a feasibility study in primates. *Neurol Res* 2003;25(8):794–800. [PubMed: 14669521]
- Qaqish BF. A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika* 2003;90(2):455–463.
- Raudenbush, SW.; Bryk, AS. Hierarchical Linear Models: Applications and Data Analysis Methods. Vol. 2. Thousand Oaks, CA: Sage Publications; 2002. p. 75-80.
- Rothman, KJ.; Greenland, S. Modern Epidemiology. Vol. 2. Philadelphia, PA: Lippincott-Raven Publishers; 1998. p. 184-188.
- Teskey GC, Flynn C, Goertzen CD, Monfils MH, Young NA. Cortical stimulation improves skilled forelimb use following a focal ischemic infarct in the rat. *Neurol Res* 2003;25(8):801–810. [PubMed: 14669522]

Wolf SL, Thompson PA, Morris DM, et al. The EXCITE trial: attributes of the Wolf Motor Function Test in Patients with subacute stroke. *Neurorehabil Neural Repair* 2005;19(3):194–205. [PubMed: 16093410]

Table 1
Primary and Secondary Outcomes and Treatment Parameters

Model	Project	Outcomes for Efficacy (E) and Persistence (P)	
		Years 1–2	Years 3–4
Rodent	Project 1: Optimize Cortical Stimulation Parameters	<p>Primary: reaching success on last day of treatment (E)</p> <p>Secondary: reaching success on any other day of testing during treatment period (E); slope of recovery curve (E)</p>	<p>Primary: ratio post-/pre-stroke reaching success on last day of treatment (E) reaching success at 1, 3, and 6 mo post-treatment (P)</p> <p>Secondary: reaching success on any other day of testing during treatment period (E); slope of recovery curve (E); correlation of reaching success and map area/map threshold (E)</p>
	Project 2^a: Optimize Rehabilitative Therapy Parameters	<p>Primary: reaching success on last day of treatment (E); reaching success at 6 mo post-treatment (P)</p> <p>Secondary: reaching success on any other day of testing during treatment period (E); slope of recovery curve (E)</p>	<p>Primary: paw misplacements on last day of treatment (E); reaching success on last day of treatment (E); reaching success at 6 mo post treatment (P)</p> <p>Secondary: paw misplacements on any other day of testing during treatment period (E); reaching success on any other day of testing during treatment period (E); slope of recovery curve (E); errors per step (E); motor performance score (E); asymmetry cylinder task (E); ipsilateral response (E)</p>
Primate	Project 3: Generalize Cortical Stimulation and Rehabilitative Therapy Parameters	<p>Primary: avg performance time on last day of treatment (E); avg performance time at 12 wks post-treatment (P)</p> <p>Secondary: avg performance time on any other day of testing during treatment period (E); slope of recovery curve (E); reaching success on last day of treatment (E); days to criterion performance (E)</p>	<p>Primary: avg performance time on last day of treatment (E)</p> <p>Secondary: avg performance time on any other day of testing during treatment period (E); slope of recovery curve (E); reaching success on last day of treatment (E); days to criterion performance (E)</p>

**Milestone
Decision to
Proceed to
Years 3–4**

^aTwo primary efficacy outcomes are given in years 3–4 because there are two studies for this project during that period.

Table 2Algorithm for Determining Efficacy/Persistence^a

Secondary Outcomes	Primary Outcome		
	$P \leq 0.05$	$0.05 < P \leq 0.2$	$P > 0.2$
2+ with $P \leq 0.05$	Shaded regions represent treatments determined to be efficacious/persistent		
1 with $P \leq 0.2$			
All with $P > 0.2$			

^a Shaded regions represent treatments determined to be efficacious/persistent

Table 3
Milestone Decision Algorithm Simulation Type I Error Results

Significance Level ^a	Near-significance Level ^a	Correlation between Tests Results ^b	Simulated Typed I Error
0.05	0.10	0.00	0.1058
0.05	0.10	0.25	0.1011
0.05	0.10	0.50	0.0819
0.05	0.10	0.75	0.0639
0.05	0.10	1.00	0.0534
0.05	0.15	0.00	0.1338
0.05	0.15	0.25	0.1299
0.05	0.15	0.50	0.1032
0.05	0.15	0.75	0.0923
0.05	0.15	1.00	0.0811
0.05	0.20	0.00	0.1581
0.05	0.20	0.25	0.1439
0.05	0.20	0.50	0.1359
0.05	0.20	0.75	0.1256
0.05	0.20	1.00	0.1095
0.05	0.25	0.00	0.1808
0.05	0.25	0.25	0.1738
0.05	0.25	0.50	0.1623
0.05	0.25	0.75	0.1454
0.05	0.25	1.00	0.1356

^aTwo-sided

^bExchangeable correlation between tests results within a project