Practical Model-Based Dose Finding in Early-Phase Clinical Trials

Optimizing Tissue Plasminogen Activator Dose for Treatment of Ischemic Stroke in Children

Harry T. Whelan, MD; John D. Cook, PhD; Catherine M. Amlie-Lefond, MD;
Collin A. Hovinga, PharmD; Anthony K. Chan, MD; Rebecca N. Ichord, MD;
Gabrielle A. deVeber, MD; Peter F. Thall, PhD

Background and Purpose—A safe and effective tissue plasminogen activator (tPA) dose for childhood stroke has not been established. This article describes a Bayesian outcome-adaptive method for determining the best dose of an experimental agent and explains how this method was used to design a dose-finding trial for tPA in childhood.

Methods—The method assigns doses to successive cohorts of patients on the basis of each dose's desirability, quantified in terms of the tradeoff between efficacy and toxicity. The tradeoff function is constructed from several pairs of equally desirable (efficacy, toxicity) probabilities specified by the physicians planning the trial. Each cohort's dose is chosen adaptively, based on dose-outcome data from the patients treated previously in the trial, to optimize the efficacy-toxicity tradeoff. Application of the method to design the tPA trial is described, including a computer simulation study to establish design properties. A hypothetical cohort-by-cohort example is given to illustrate how the method works during trial conduct.

Results and Conclusions—Because only a dose that is both safe and efficacious may be selected and the method combines phase I and phase II by integrating efficacy and toxicity to choose doses, it avoids the more time-consuming and expensive conventional approach of conducting a phase I trial based on toxicity alone followed by a phase II trial based on efficacy alone. This is especially useful in settings with low accrual rates, such as trials of tPA for pediatric acute ischemic stroke. (*Stroke*. 2008;39:2627-2636.)

Key Words: stroke ■ tPA ■ dose finding ■ Bayesian statistics ■ phase I/II trial

Every year, at least 6 in every 100 000 children under the age of 18 years have a stroke.[1,2] Ten percent of these children die, 20% have another stroke, and 70% have seizures or other neurologic deficits.[3] The attendant healthcare needs of these children can persist for many decades and result in the loss of the most productive years of life.[4]

When administered acutely to appropriate adults, intravenous (IV) tissue plasminogen inhibitor (tPA) is relatively safe for treating ischemic stroke, but it is associated with a 6.4% risk of symptomatic intracerebral hemorrhage (SICH).[5,6] Although children often arrive at hospital within the time window for tPA treatment,[7,8] age-appropriate tPA safety data and dosing guidelines are lacking. There are critical physiologic differences between the hemostatic systems of children compared with those of adults,[9,10] including decreased levels of many coagulants, suggesting that the dosing guidelines for

tPA for stroke in adults may not be safe or efficacious in children. In children, the fibrinolytic system is overall hypoactive. By the age of 2 years, plasminogen concentrations are compatible to those of an adult. Baseline tPA concentrations in the blood of 1- to 16-year-old children are ≈50% lower than in adults.[9] A study of 14- to 18-year-old female patients demonstrated that teenagers have lower veno-occlusive–stimulated fibrinolytic activity (50% to 70%) compared with that in adult men and women.[11]

Andrew et al[9] showed that in teenagers, concentrations of plasminogen activator inhibitor-1, the predominant inhibitor of tPAs, were increased compared with those observed in adults. Plasminogen and its inactivating protein (α_2-antiplasmin) are not similar across childhood and in the adult.[9] The need for an increased tPA dose in children to promote fibrinolysis is supported by the lower baseline levels

of tPA (lower activity) and the increased plasminogen activator inhibitor concentrations (increased level of inhibition). Another factor suggesting that increased doses of tPA are needed in children is related to differences in the pharmacokinetics of tPA in children and adults, which suggest that higher tPA doses (mg/kg body weight) are needed to achieve comparable plasma concentrations in children because of an increase in the volume of distribution and possibly a more rapid hepatic clearance. In adults, tPA distributes in a volume approximating plasma volume which, when normalized to weight, is much higher in children.[12] Therefore, developmental differences in hemostasis and drug distribution suggest that safety and efficacy data cannot be used to extrapolate tPA dosages from adults to children.

tPA is increasingly being given for childhood stroke in the absence of safety and efficacy data for children and often outside adult standards, including intervals from stroke onset well beyond 3 hours.[13,14] Early-phase clinical trials are therefore urgently needed to establish whether tPA is effective in childhood stroke with acceptable levels of toxicity.

Dose-Finding Clinical Trial Objective

This article has 2 purposes. The first is to describe a Bayesian statistical procedure, proposed by Thall and Cook,[15] for determining the best dose of an experimental agent. The second purpose is to explain how this method was used in the design of a dose-finding trial for tPA in childhood acute ischemic stroke (AIS). Section 2 provides general reviews of Bayesian statistics and outcome-adaptive dose selection. Section 3 reviews the dose-finding method, including the underlying dose-outcome probability model, efficacy-toxicity tradeoffs, and the adaptive algorithm for assigning doses to patients during trial conduct. In section 4, we explain how the method was applied to design the trial of tPA for treatment of AIS in children. This example includes computer simulations to describe the design's average behavior, and we also give a hypothetical cohort-by-cohort illustration of how the method works in practice. We close with a discussion of practical and ethical issues that should be considered when implementing the method.

Statistical Preliminaries

Bayesian Statistics

Because the dose-finding method that we describe relies on a Bayesian statistical framework, we first briefly review the Bayesian paradigm. In Bayesian statistics, there are 2 types of objects. The first is a vector of 1 or more parameters, which we denote by the Greek symbol θ, and the second is the observed data. Although parameters are not observed, they describe important aspects of the phenomenon giving rise to the data. For example, θ may include the probabilities of efficacy and toxicity in a dose-finding trial, the effects of patient covariates such as age or disease severity, or median survival time. A Bayesian model has 2 components. The first is a function, *likelihood* (data | θ), which describes the probability of the observable data for a given θ. Some common likelihood functions are the normal (bell shaped) distribution, the binomial distribution for binary (success/failure) data, and

the exponential distribution for event times. In contrast with conventional frequentist statistics, which treats θ as fixed but unknown, the Bayesian paradigm considers θ to be random. Thus, a Bayesian model also requires a probability distribution, *prior* (θ), which describes what is known about θ before observing the data. So-called "uninformative" *priors* often are used in settings where little is known about θ before the data are observed, and such *priors* often are formulated to contain the amount of information in 1 or 2 observations. In contrast, when historical data are available or the investigators have substantive knowledge about θ, a so-called "informative" *prior* may be formulated to reflect the historical data or prior knowledge. Once *prior* (θ) is established, the data are then used to learn about the probability distribution of θ by applying Bayes' Theorem, which combines the *prior* and the *likelihood* to compute the *posterior* distribution,

$$posterior\ (\theta \mid data) \propto likelihood\ (data \mid \theta) \times prior\ (\theta)$$

The *posterior* describes what one knows about θ after observing the data. The symbol "$\propto$" means "is proportional to" and reflects the technical requirement that the product *likelihood* (data | θ) $\times$ *prior* (θ) must be divided by p (data) to obtain the *posterior*. Thus, Bayes' Theorem incorporates the information in the data by turning one's *prior* into a *posterior*, which is then used to compute probability statements about θ and make statistical inferences.

Bayes' Theorem may be applied repeatedly in successive stages as new data become available by using the *posterior* obtained after each stage as the *prior* for the next stage. By performing the *prior*-to-*posterior* computation repeatedly and making a decision or taking an action based on quantities computed from the *posterior* at each stage of this process, the Bayesian paradigm provides a natural framework for learning and taking actions on the basis of accumulating data during a clinical trial. Bayesian adaptive dose finding is an example of this type of dynamic process. A recent commentary on Bayesian methods is given by Berry.[16]

Outcome-Adaptive Clinical Trials

Outcome-adaptive statistical methods in clinical trials repeatedly use the data from patients who have been treated previously in the trial to make interim decisions within the same trial. The data generally consist of each patient's treatment and outcome, and some methods also use patient prognostic covariates. Examples of interim decisions include whether to stop or continue the trial, what to conclude if the trial is stopped, whether to terminate a particular treatment arm, revision of the planned sample size, or what treatment or dose to assign to the next patient or cohort of patients. Outcome-adaptive methods contrast sharply with the common statistical practice of leaving information that accrues during a trial untouched throughout the course of the study and only analyzing it at the end.

An outcome-adaptive dose-finding procedure starts by treating the first cohort at a dose that initially is considered to be acceptably safe. Usually, this is the lowest dose among those being studied, although this is not necessarily the case. Most outcome-adaptive phase I trials use a binary indicator of

toxicity as the outcome. When the outcomes from the first cohort of patients treated at the initial dose are observed, these data are used as a basis for choosing the dose for the next cohort. This process is repeated, with each new cohort's dose chosen on the basis of the previous patients' dose-toxicity data. Although there are many methods for dose finding based on toxicity in phase I clinical trials, by far the most commonly used are variants of conventional "3+3" algorithms. Although 3+3 methods are outcome-adaptive, they only use the data from the most recent 1 or 2 cohorts. Numerous computer simulation studies have shown that 3+3 algorithms have very poor properties when compared with outcome-adaptive methods that use all of the currently available data.[17,18]

Bayesian outcome-adaptive methods base each interim decision on the most recently updated *posterior* (see equation above), and thus use all of the currently available data for each decision. For each new cohort of patients, the most recent data are used to update the dose-outcome probabilities by using the new *posterior*, and this is used as a basis for choosing the new cohort's dose. This *posterior* then becomes the new *prior*, and the process is repeated each time new data become available and a dose must be chosen. Each new dose may be higher than, the same as, or lower than the previous dose, depending on the data.

Review of the Method

The dose-finding method described herein differs from conventional phase I methods in 2 essential ways. First, it uses both efficacy (E) and toxicity (T) to determine doses for successive patient cohorts rather than relying on T alone. The second difference is that it is based on a Bayesian statistical framework. The method has 3 basic components. The first is a Bayesian model that accounts for the probabilities of E, often referred to as "response," and T as functions of dose. The second component consists of criteria that allow the investigators to determine the set of doses that have both acceptably low toxicity and acceptably high efficacy. The third component is a function for evaluating the tradeoff between the probabilities of E and T for each dose, and it provides a basis for quantifying the desirability of each dose. The tradeoff function is based on several elicited (E, T) probability pairs that are considered by the physician to be equally desirable targets. Each cohort receives the most desirable dose based on the most recently updated *posterior*. The method may be called a "phase I/II" design because it combines the goals of conventional phase I and phase II trials, including evaluating toxicity, evaluating efficacy, and finding an acceptable dose in affected subjects.

Determining a Starting Dose

The starting dose is not necessarily the lowest dose being considered. When specifying the doses to be studied, a physician may initially specify a set of doses with the lowest as the starting dose. However, with this approach, if the lowest dose is found to be unacceptably toxic, the trial must be stopped. It is common practice to then add 1 or more lower doses and then restart the trial, subject to institutional review board approval. To deal with such eventualities ahead of time, it is very useful to specify such lower doses initially so that they may be included in the design from the start, to avoid having to stop the trial, add lower doses, and then restart. If the trial begins at the same initial dose originally specified, the starting dose is now no longer the lowest. Phase I designs that use toxicity as the only criterion for dose finding often start at the lowest dose for fear of excessive toxicity. However, one may argue that dose-finding trials that account for efficacy should start at the highest dose for fear of administering an ineffective dose. In practice, a dose-finding trial that uses efficacy may start at the highest dose for this reason. Thus, when doing dose finding based on both efficacy and toxicity, choosing a starting dose depends on one's prior relative concern of overdosing and underdosing patients.

Outcomes and Probabilities

In practice, the definitions of "efficacy" and "toxicity" will vary widely, because these events are highly dependent on the particular medical setting. Consequently, it is essential that each outcome be defined collaboratively by the researcher and statistician in a manner that is appropriate to the particular trial at hand. Likewise, the probabilities of these events that are considered acceptable also will depend on the particular disease being treated, the trial's entry criteria, and the rates of efficacy and toxicity that may be expected with whatever standard therapies may be available. Whichever outcome variable is chosen, it should be well defined, clearly measurable, and clinically relevant; because the design requires outcomes that can be observed quickly enough to choose doses adaptively, the efficacy outcome should have a reasonable association with a long-term treatment benefit.

In many settings, there is a positive association between toxicity and efficacy, which provides the rationale for the conventional practice of using toxicity alone to identify a "maximally tolerated dose" when evaluating cytotoxic agents. When toxicity and efficacy occur independently or are negatively associated, which may be the case with biologic agents having complex effects on clinical outcome, conventional dose-finding methods may be inadequate or just plain wrong. In AIS, for a fibrinolytic agent, efficacy can be defined as the presence of recanalization because it constitutes reversal of the ischemic occlusion.

In AIS, the toxicity of a fibrinolytic agent is defined as significant hemorrhage manifesting as either an SICH or a symptomatic systemic bleed. Using this as a combined outcome variable, though intuitively clear, may still pose some difficulty in terms of interpretation. It may be assumed that SICH is associated with recanalization or efficacy, as arterial blood flow is required to create a significant hemorrhage. Nonetheless, it is unclear that this is necessarily the case. Hemorrhagic transformation as opposed to SICH, which occurs in 15% to 43% of patients with AIS, is thought by some to be due to augmented collateral circulation, and timing rather than recanalization may be the salient risk factor for hemorrhagic transformation.[19] tPA may also have excitotoxic or cytotoxic effects that contribute to the risk of hemorrhagic transformation and SICH.[20] Therefore, it is unclear whether toxicity and efficacy have a positive associ-

ation, have a negative association, or are independent of each other.

The dose-finding method described herein provides a compromise between the scientific goal of the study, which is to determine the most appropriate dose for future patients as reliably as possible, and the ethical goal of giving each successive cohort treated during the trial the best possible dose based on the most recent interim data. Early in the trial, the successively chosen doses may vary substantially, because the amount of data is small. As the data accumulate and each successive *posterior* becomes more informative, the method is more likely to assign patients to the most desirable dose or doses, if they exist, and the emphasis shifts toward reducing variability in outcomes. Practically, this can be thought of as affirming or confirming the results at or near the final chosen dose with a larger, more focused sample.

To apply the method, E and T may be defined in 2 different ways. In the first, E and T are defined in such a way that both cannot occur, so the 3 possible outcomes are {E, T, neither}. The second way, which was used in the tPA trial design, allows the possibility that both E and T may occur in the same patient. Thus, each patient has 1 of 4 possible elementary outcomes: {E and no T}, {E and T}, {no E and no T}, and {no E and T}. The event E thus can occur in 2 different ways: {E and T} or {E and no T}, and similarly, T occurs if the patient's outcome is either {E and T} or {no E and T}. A patient with outcome {E and T} has achieved the desired efficacy event but also suffered toxicity, so this is at once a good outcome and a bad outcome. Although one may hope to achieve the best possible outcome, {E and no T}, and avoid the worst possible outcome, {no E and T}, in practice one should account for all 4 possibilities. To sort out this structure and provide a method for choosing doses that makes sense scientifically and ethically, the method relies on statistical models for the probability of E as a function of dose, which we denote for brevity by $\pi_E(dose)$, and the probability of T as a function of dose, denoted by $\pi_T(dose)$. The latter represents the usual dose-toxicity curve used by conventional methods based on toxicity alone. The algorithm for choosing doses is built around the probability pair $\pi(dose)=\{\pi_E(dose), \pi_T(dose)\}$. The method may be tailored to accommodate trials wherein the definition of E excludes the occurrence of T, and thus, only the 3 elementary outcomes, E, T, and {neither E nor T}, are possible. Technical details are given in Thall and Cook[15] and Thall et al.[21]

Acceptability Criteria

The method uses 2 different types of criteria to choose doses, both computed with the *posterior* based on the most current data under the Bayesian model. The first criterion determines whether each dose is acceptable. Let π_E be a fixed lower limit on $\pi_E(dose)$ and π_T, a fixed upper limit on $\pi_T(dose)$. The particular numeric values of π_E and π_T should be motivated by the definitions of E and T and what are considered medically acceptable levels of these events when treating the particular disease. Selection of these values can be guided on the basis of existing clinical or published data based on an established standard (eg, other treatment, no treatment, or heparin) or the natural rate of an event (rate of spontaneous

efficacy-recanalization or toxicity-SICH) occurring in the studied population. For example, if toxicity includes regimen-related death and this has a historical rate of 10% with standard treatment, whereas only doses with at least a 50% efficacy rate are of interest, then $\pi_T=0.10$ and $\pi_E=0.50$ are appropriate. In contrast, for a combined phase I dose finding and phase IIA activity trial of a new biologic anticancer agent that does not carry the risk of regimen-related death but at worst has moderate toxicities, with a targeted 20% or larger tumor-response rate, $\pi_T=0.30$ and $\pi_E=0.20$ may be appropriate Given the current data, a dose is acceptable if it has both acceptably high efficacy and acceptably low toxicity, formally, if it is not unlikely that either $\pi_E(dose)$ is below π_E or that $\pi_T(dose)$ is above π_T. These 2 criteria act as gatekeepers, 1 for E and 1 for T. Only acceptable doses may be used to treat patients in the trial. If there are 2 or more acceptable doses, however, then an additional criterion is needed to select the best among them.

Dose Desirability and Efficacy-Toxicity Tradeoffs

Because the probability pair $\pi(dose)=\{\pi_E(dose), \pi_T(dose)\}$ for each dose is 2-dimensional, to use $\pi(dose)$ as a basis for selecting a best acceptable dose it must be reduced to a 1-dimensional value. This is done by using the following geometric construction. First, the researchers must specify 3 or more probability pairs (ie, probability of E and T) that they consider equally desirable targets. These are represented by the triangular points in Figure 1. Each elicited target represents a tradeoff between the probability of achieving a response and the risk of toxicity. A curve is fit to the pairs, and this curve is called the target E-T tradeoff contour. The reference contour is used to generate a family of tradeoff contours so that every possible π falls on exactly 1 contour. A numeric desirability is assigned to each contour, so that all π on the same contour are equally desirable. Figure 1
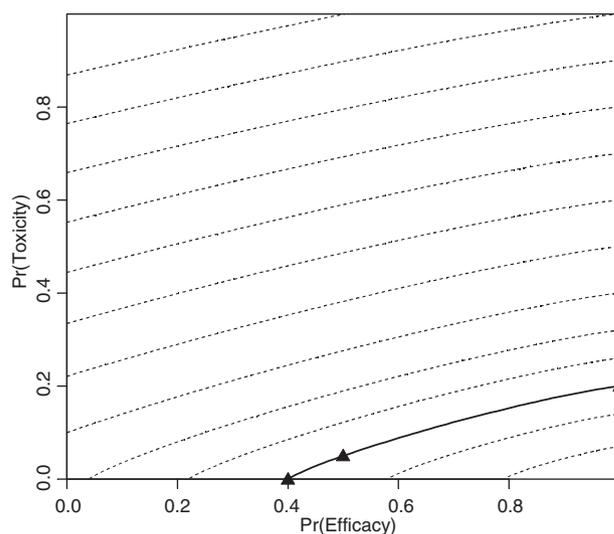
Figure 1. E–T tradeoff contours. The 3 elicited, equally desirable target $\pi=\{Pr(Efficacy), Pr(Toxicity)\}$ pairs are represented by triangles, and the solid curve is the resulting target tradeoff contour. Each dashed line is a tradeoff contour on which all π pairs are equally desirable. The numeric desirability equals 1 for all π pairs on the target contour, is <1 for all π pairs above the target contour, and is >1 for all π pairs below the target contour.

illustrates the contours used for the tPA trial, with the target contour given by the thick line. In Figure 1, all π on the target contour have desirability equal to 1, all π on contours above the target have desirability <1, with the desirability decreasing as π moves away from the target contour toward the worst possible point $\pi=(0,1)$. The desirability of each point below the target contour is >1 and increases as π moves toward the ideal point $\pi=(1,0)$, which corresponds to certain response and no risk of toxicity. During the trial, the current *posterior* mean of π(dose) is computed for each acceptable dose, the contour where this pair is located is determined, and the desirability of that contour is then assigned to d. The acceptable dose having the highest desirability is assigned to the next cohort. Technical details of this construction are given in Thall et al[21] and Thall.[22]

Trial Design and Conduct

To construct a design, one must first establish trial inclusion/exclusion criteria, treatment and doses (number of doses and actual dose amounts), the definitions of E and T, maximum sample size, and cohort size. The *prior* is constructed on the basis of *prior* means of π_E(dose) and π_T(dose) at each dose to be studied in the trial. The acceptability boundaries π_E and π_T must be specified, as well as target π pairs to construct the target contour and the resulting family of tradeoff contours.

The rules for trial conduct are as follows: (1) Treat the first cohort at the starting dose specified by the researchers. (2) For each cohort after the first, if there is at least 1 acceptable dose, then treat the next cohort with the most desirable acceptable dose, subject to rules 3 and 4. (3) For each cohort after the first, no untried dose may be skipped, either when escalating or de-escalating. (4) At any interim point in the trial, if there are no acceptable doses, stop the trial and do not select any dose. (5) If the trial is not stopped early and there is at least 1 acceptable dose at the end, then select the acceptable dose having the largest desirability.

Rule 4 may be regarded as a combination of 2 more conventional rules, a "safety" rule that stops a trial if the treatment is too toxic and a "futility" rule that stops a trial if the treatment is ineffective. Consequently, the numeric values of the limits π_E and π_T and the cutoffs p_E and p_T used to define the 2 "gatekeeper" criteria are very important.

Computer Simulation as a Design Tool

Because this design is complex, before using it to conduct a trial, it is essential to simulate the trial on the computer under each of several dose-outcome scenarios to evaluate the design's operating characteristics (OCs). A dose-outcome scenario consists of fixed values of π_E(dose) and π_T(dose) for each dose to be studied in the trial. With the design, the trial is simulated a large number of times (1000 or more) under each scenario, and the results are recorded. The OCs consist of the selection probabilities and average sample sizes at each dose and the probability of stopping the trial early. These values are analogous to the usual type I error probability and power of a conventional test of hypothesis. The simulation results may be used to study the design and, if necessary, adjust its parameters to obtain good OCs. An acceptable design must have a high probability of stopping early and

choosing no dose in scenarios wherein all doses are unacceptably toxic or ineffective and reasonably high probabilities of selecting desirable doses when they exist. Computer simulation allows one to conduct a "thought experiment" ahead of time without risking patient's lives by using preliminary simulation results as a tool to calibrate the design.[15–22] As an additional check, it also is useful to see whether the design behaves reasonably at the start for specified data from the first 1 or 2 cohorts. A computer program, EffTox, which performs all necessary computations, is freely available for download at the website http://biostatistics.mdanderson.org/SoftwareDownload/.

EffTox requires numeric values of all of the quantities noted earlier, and it computes the *prior* and the family of tradeoff contours. EffTox includes a graphical user interface for plotting the target points and the resulting target contour during the elicitation process, so that one may modify the targets interactively. The targets may be established at the same time as the anticipated mean π_E(dose) and π_T(dose) values used to construct the *prior*. The targets represent what one would like to achieve, similar to specifying an alternative parameter value when constructing a test of hypotheses. In contrast, the elicited *prior* means represent what one anticipates actually will happen, and they essentially provide a starting point for the model. EffTox also supports simulations and trial conduct.

Trial of tPA for Pediatric AIS

Rationale for Optimizing tPA Dose in Children

Because AIS in childhood is significantly less common than adult stroke and has significant delays to diagnosis,[8] identification of eligible study subjects within the acute time period for tPA (3 hours from stroke onset for IV therapy and 6 hours from onset for intra-arterial therapy) may be difficult. Because traditional phase I clinical trials ignore efficacy and thus are limited in their ability to fulfill these criteria, a phase I/II trial was designed in which tPA would be given to patients age 2 to 17 years with radiographically confirmed AIS or cerebral artery occlusion, IV if within 3 hours or intra-arterially if within 6 hours of the onset of symptoms. For simplicity, the intra-arterial portion of the study has been excluded from the dose-finding design discussed herein.

The outcomes used by the dose-finding method are evaluated 48 hours after beginning treatment. Efficacy is defined for the purposes of this trial as angiographic recanalization or restoration of flow past the area of occlusion on follow-up magnetic resonance angiography. As with all short-term outcomes used to characterize treatment effect for the purpose of making outcome-adaptive decisions, the method relies on the assumption that efficacy as defined is associated with long-term treatment benefit. As in any early-phase trial, although efficacy certainly is not a perfect surrogate for long-term treatment effect, this assumption is a reasonable compromise made to construct a feasible design. Toxicity is defined as fatal or symptomatic intracranial or systemic hemorrhage (ICH). SICH is defined as a newly identified hemorrhage seen on neuroimaging associated with a worsening of 4 or more points on the Peds-National Institutes of

	tPA Dose, mg/kg				
	0	0.6	0.8	1.0	1.2
Efficacy	0.15	0.50	0.55	0.65	0.70
Toxicity	0.01	0.04	0.07	0.15	0.20

For each tPA dose, the tabled values are the mean probabilities of E and T that were anticipated, based on clinical experience and interpretation of published results, before trial conduct.

Health Stroke Scale or a change in the level of consciousness. Typically, the occurrence of any ICH with tPA administration is considered dose-limiting. However, there may be sequelae of ICH thought to be asymptomatic in the acute period. Thus, given the prognosis of the eligible patients, we think it appropriate to use a less-restrictive definition of toxicity.

tPA Trial Design

The dose-finding method was applied as follows. A maximum of 24 patients will be treated in cohorts of size 2, for a maximum of 12 cohorts. Each cohort will receive treatment with IV tPA chosen from the established range of 4 possible doses (0.6, 0.8, 1.0, or 1.2 mg/kg body weight), with the first cohort treated at 0.8 mg/kg. Candidates with strokes who missed the established 3-hour cutoff window for IV tPA treatment but are imaged in the first 6 hours after symptom onset will serve as a parallel nonrandomized control group for

the protocol, because they will be treated at 0 mg/kg tPA. Based on the same rationale that motivated the definitions of E and T, the acceptability limits $\pi_E = 0.20$ and $\pi_T = 0.20$ will be applied. The target tradeoff contour is based on the 3 equally desirable targeted tradeoff probability pairs (0.50, 0.05), (0.40, 0), and (1.0, 0.20). In Figure 1, the resulting target E-T tradeoff contour is illustrated by the solid curve, and other contours are given by dashed lines. Table 1 gives the means of π_E(dose) and π_T(dose) used to determine the *prior*, which was calibrated to contain very little information, with the effective sample size between 1.34 and 1.50 for the *prior* on each π_E(dose) or π_T(dose).

On the basis of the ECAS I trial that reported 19% SICH, we adopted this as an acceptable upper limit of the toxicity of our trial. However, a relative high efficacy is necessary to justify such a relative high toxicity, and thus the limit of efficacy is set at 70% with the premise that 100% efficacy cannot be achieved. Furthermore, a recent meta-analysis showed a recanalization rate as high as 70%[23,24] and efficacy in systemic thrombolysis in children for non–central nervous system thrombosis being close to 70%.[25]

Computer Simulation Results for the tPA Trial Design

The OCs of this trial design were computed under 8 dose-toxicity scenarios, with maximum sample sizes of 24 or 48 and cohorts of size 2 or 3, for a total of 32 combinations of

Table 2. Computer Simulation Results for the tPA Trial Conducted With the E–T Dose-Finding Design, Under Each of 4 Dose-Outcome Scenarios

	tPA Dose, mg/kg				
Scenario	0.6	0.8	1.0	1.2	None
1					
True π_T, π_E	0.04, 0.50	0.05, 0.62	0.15, 0.67	0.20, 0.70	
Desirability	1.04	1.21	0.84	0.70	
% Selected	41	46	11	1	0
No. of patients treated	9.7	9.8	3.9	0.6	
2					
True π_T, π_E	0.10, 0.30	0.20, 0.40	0.40, 0.45	0.50, 0.50	
Desirability	0.59	0.45	0.19	0.13	
% Selected	77	16	1	0	7
No. of patients treated	15.7	5.6	1.3	0.2	
3					
True π_T, π_E	0.01, 0.20	0.02, 0.40	0.05, 0.60	0.10, 0.40	
Desirability	0.70	0.94	1.17	0.70	
% Selected	11	28	40	21	0
No. of patients treated	4.6	6.7	7.7	4.5	
4					
True π_T, π_E	0.40, 0.05	0.42, 0.08	0.45, 0.10	0.50, 0.20	
Desirability	0.11	0.10	0.09	0.08	
% Selected	1	4	2	1	92
No. of patients treated	4.2	1.9	0.8	0.2	

Each scenario is characterized by the tabled "true π_T, π_E" values, which were used to simulate patient outcomes under that scenario according to the dose-finding design. For each dose, "desirability" quantifies the true π_T, π_E pair under the E–T tradeoff (Figure 1) used by the design. Each tabled "% selected" and "No. of patients treated" at each dose is the average value over 1000 simulated trials under the assumed scenario.
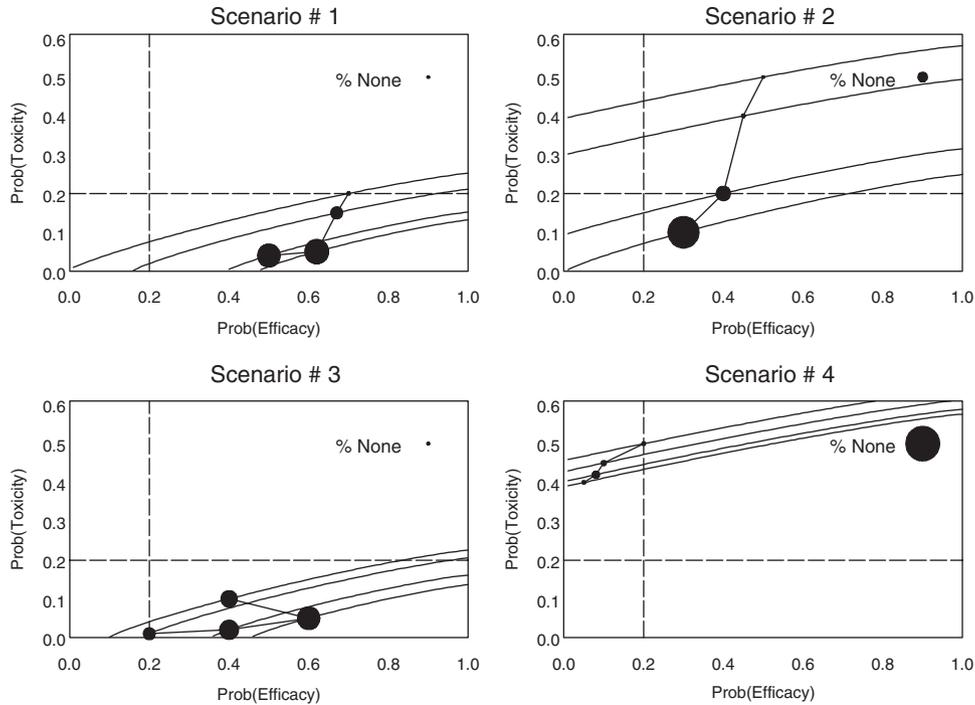
Figure 2. Graphic illustration of the selection probability and desirability of each dose in the simulation results summarized in Table 2. For each dose, the contour where the *posterior* mean of {Pr(Efficacy), Pr(Toxicity)} is located is given to illustrate the comparative desirabilities of the 4 doses based on the observed data. The selection probabilities of doses 1, 2, 3, and 4 and the probability of selecting no dose are represented by the areas of the corresponding discs.

scenario and design. For illustration, we present computer simulation results for the final design under each of 4 dose-outcome scenarios. The results are summarized in Table 2 and Figure 2. For each scenario in the figure, the true values of $\pi_E(dose)$ are given on the horizontal axis and $\pi_T(dose)$ on the vertical axis. For each dose of 0.6, 0.8, 1.0, and 1.2 mg/kg, a shaded circle having area equal to the dose's selection probability is given at the true $\pi(dose) = \{\pi_E(dose), \pi_T(dose)\}$ location, and the tradeoff contour passing through $\pi(dose)$ is shown as a solid curve. The admissibility limits $\pi_E = 0.20$ and

$\pi_T = 0.20$ are given as dashed straight lines in each plot. For example, the upper left-hand plot in Figure 2 illustrates the simulation results for scenario 1 in Table 2, where the 4 doses have respective desirability values of 1.04, 1.21, 0.84, and 0.70 and are selected with percentages 41, 46, 11, and 1. In contrast with scenario 1, in which all doses are safe, under scenario 2 the toxicity probabilities $\pi_T(1.0) = 0.40$ and $\pi_T(1.2) = 0.50$ are well above the acceptability limit $\pi_T = 0.20$, and the method recognizes this. Scenario 3 is a difficult case in which all doses are safe, but $\pi_E(dose)$ increases from

Table 3.   Hypothetical Case-by Case Illustration of the tPA Trial Design for a 12-Cohort Trial With 2 Patients per Cohort

| Cohort		Outcomes				tPA Dose, mg/kg			
		No E and T	No E and No T	E and T	E and No T	0.6	0.8	1.0	1.2
	Assigned Dose					*Posterior* Desirability of Each Dose			
1	0.8	0	1	0	1	0.88	**1.17**	1.26	0.97
2	1.0	0	2	0	0	**0.83**	0.78	0.79	0.70
3	0.6	0	2	0	0	0.63	0.68	**0.78**	0.76
4	1.0	0	0	0	2	0.64	0.85	1.26	**1.37**
5	1.2	0	0	1	1	0.61	**0.85**	1.20	0.44
6	1.0	0	2	0	0	0.60	0.74	**1.00**	0.44
7	1.0	1	0	1	0	0.58	**0.66**	0.52	0.11
8	0.8	0	2	0	1	0.56	**0.63**	0.52	0.12
9	0.8	0	1	0	1	0.59	**0.70**	0.56	0.11
10	0.8	0	1	0	1	0.61	**0.75**	0.60	0.11
11	0.8	0	0	1	1	0.67	**0.75**	0.46	0.13
12	0.8	0	2	0	0	0.76	**0.86**	0.49	0.12

The largest desirability at each stage is given in boldface type.
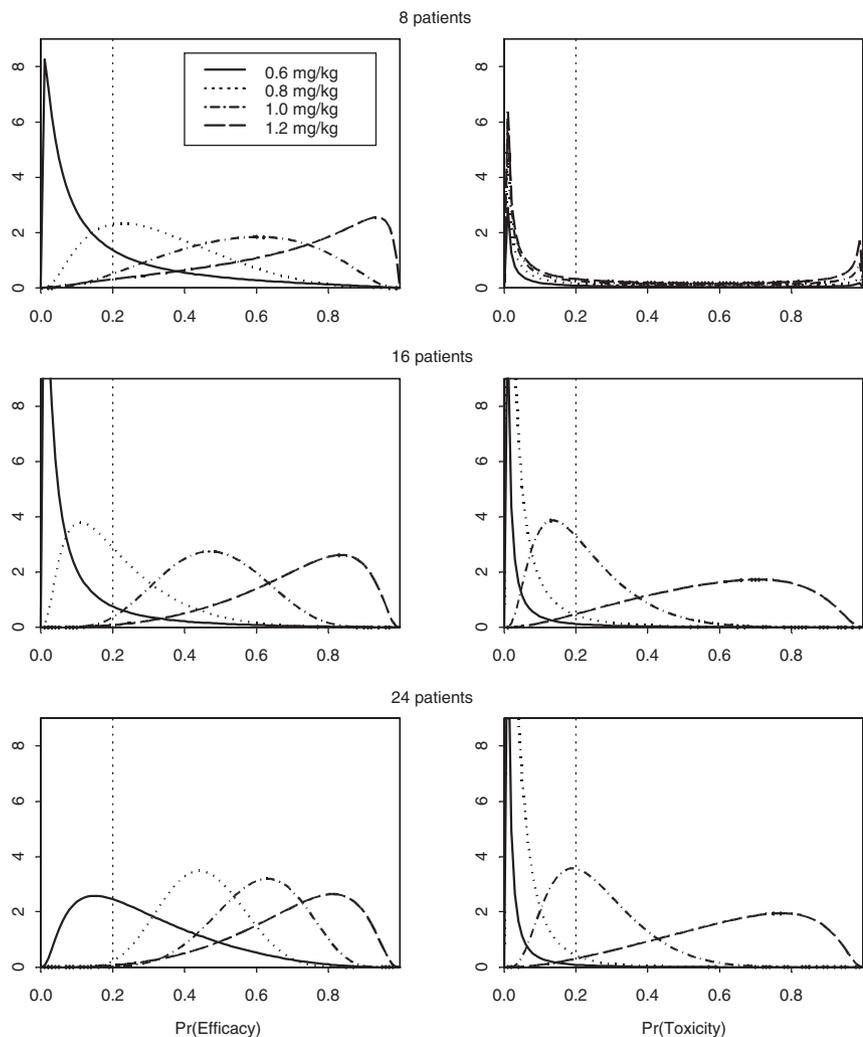
Figure 3. In the case-by-case example, *posterior* distributions of π_E(dose) and π_T(dose) for each dose=0.6, 0.8, 1.0, and 1.2 mg/kg tPA, based on data from 8, 16, and 24 patients.

$\pi_E(0.8)=0.40$ to $\pi_E(1.0)=0.60$ and then decreases to $\pi_E(1.2)=0.40$; ie, efficacy is not monotone increasing with dose. The method recognizes this, selecting dose=1.0 the largest percentage of the time. Recall that the maximum sample size is only 24. With larger sample sizes, the probability of selecting the best dose in such cases also increases. Finally, scenario 4 shows that if all 4 doses are too toxic compared with the limit $\pi_T=0.20$ and the efficacy limit $\pi_E=0.20$ is only achieved at the highest dose, the method stops the trial with no dose selected 92% of the time. That is, the design is safe.

A Cohort-by-Cohort Illustration

Table 3 presents a hypothetical example of the outcomes for each of 12 successive cohorts, along with the *posterior* desirability of each dose. Recall that the dose having the highest desirability after a given cohort is assigned to the next cohort. The design starts at the second-highest dose, 0.8 mg/kg, and the first 2 patients have elementary outcomes (no E and no T) and (E and no T), so marginally there are 1/2 efficacies and 0/2 toxicities, and the *posterior* gives the highest desirability of 1.26 to dose 1.0 mg/kg. The second cohort thus is treated at this dose, both patients have outcome (no E and no T), all 4 desirabilities decrease, and the lowest

dose, 0.6 mg/kg, becomes most desirable. The 2 patients in cohort 3 thus are treated at 0.6 mg/kg, both have outcome (no E and no T), and the 4 desirables all decrease again, but now 1.0 mg/kg again becomes most desirable, by a small margin over the highest dose of 1.2 mg/kg. Thus, the 2 patients in cohort 4 are treated at 1.0 mg/kg, and both have outcome (E and no T). The desirability changes substantially, with dose 1.2 mg/kg becoming the most desirable, so cohort 5 is treated at this dose. As shown by the remainder of Table 3, the trial de-escalates back to 0.8 mg/kg and remains there for the last 5 cohorts. This example illustrates typical behavior of the method, in that it is most variable early in the trial, reflecting the facts that very little data are available and the method may escalate to a dose, de-escalate, and later re-escalate as more data become available. It also shows how the Bayesian model "learns" about the dose-toxicity and dose-efficacy curves as data from the trial become available. Of the 24 patients, only the 2 in cohort 3 were treated at 0.6 mg/kg, and only the 2 in cohort 5 were treated at 1.2 mg/kg, with the remaining 20 treated at the middle doses of 0.8 or 1.0 mg/kg. In all, 4/24 toxicities and 11/24 responses were observed. This example illustrates the inherent statistical difficulties in small-scale dose-finding trials, namely, that very often few dose levels are actually used and some have very few patients. One may
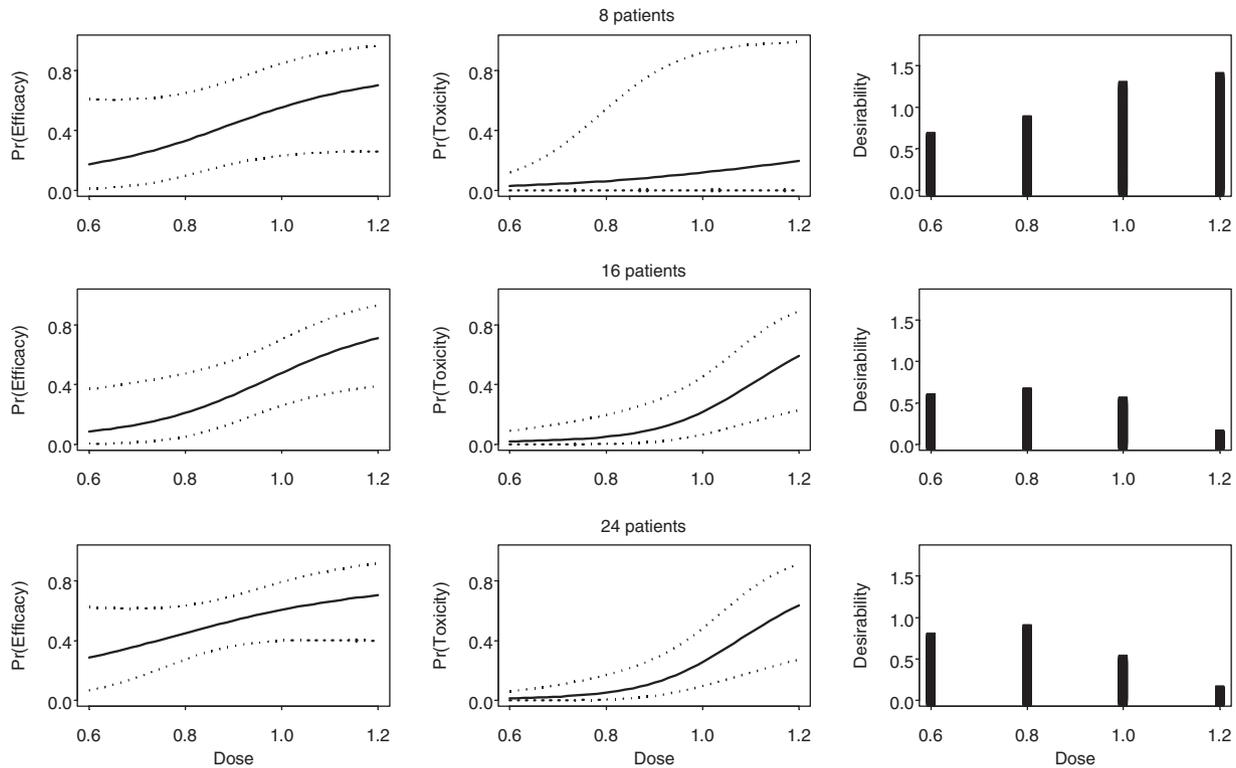
Figure 4. In the case-by-case example, *posterior* means (heavy solid lines) and 95% credible intervals (dotted lines) of Pr(Toxicity dose)=π_T(dose) (column 1), Pr(Efficacy   dose)=π_E(dose) (column 2), and desirability (column 3) as dose is varied from 0.8 to 1.2 mg/kg, based on data from 8, 16, and 24 patients.

consider the last portion of the trial, in which the final 10 patients are all treated at the same final dose level, to be its "phase II" stage. However, in general, there really is no separation between "phase I" and "phase II" with this design, and it may switch doses at any point on the basis of new data.

Based on the data in Table 3, Figure 3 gives the *posterior* distributions of π_E(dose) and π_T(dose) for each dose of 0.6, 0.8, 1.0, and 1.2 mg/kg tPA after 8, 16, and 24 patients. The *posteriors* after 8 patients show that a small amount is known about π_E(dose), with the highest 2 doses clearly acceptable relative to π_E=0.20, and all doses appear to have little risk of toxicity. The *posteriors* after 16 patients are much more informative, with the distributions of π_E(0.6), π_E(0.8), π_E(1.0), and π_E(1.2) showing a clear separation and dose 1.2 mg/kg clearly too toxic. The final *posteriors* after 24 patients show that the highest 3 doses are all acceptably efficacious with 1.2 mg/kg being the best, but only doses of 0.6 and 0.8 mg/kg are clearly safe, so on the basis of this representation, it appears that 0.8 mg/kg is best. Figure 4 summarizes the information in a different way, giving the *posterior* means (heavy type lines) and 90% credible intervals (upper 95th and lower 5th percentiles given as dotted lines) of π_T(dose) and π_E(dose) and the corresponding desirability as dose is varied from 0.8 to 1.2 mg/kg. This figure shows how one may learn about the 2 probability curves as the data accumulate, because the mean curves change and the 90% intervals become more narrow. The corresponding *posterior* mean desirability plots illustrate how the information in the *posteriors* is synthesized into a single value for each dose, with 1.0 and 1.2 mg/kg the 2 best after 8 patients, 0.8 the best by a small margin after 16

patients, and 0.8 still the best at the end of the trial but by a larger margin.

The decision to escalate to an untried dose must be based on a prediction of what is likely to occur at that higher dose. This prediction is based on an extrapolation of the fitted model for π_E(dose) and of π_T(dose) to higher values that have not yet been tried. Finally, the decisions made by the method described herein may be counterintuitive to those who are accustomed to using toxicity alone for dose finding. For example, the method is likely to stay at a safe and effective dose rather than escalate, whereas methods based on toxicity alone are likely to escalate if toxicity is under control.

Discussion

A number of other methods for dose finding with the use of outcomes that go beyond toxicity have been proposed.[26–31] The Thall-Cook method is the only 1 to formalize efficacy-toxicity tradeoffs. This methodology requires substantial input from the researcher and considerable effort to construct the design, and it also requires computer programs for both trial design and trial conduct. From a scientific or clinical perspective, however, the greater input required from the researchers may be considered a virtue, rather than a drawback, of the method. A limitation is the practical necessity for using short-term outcomes, which requires an assumed association between efficacy and long-term treatment benefit. However, this is a design for a phase I/II trial aimed at determining an optimal dose, to be used in a subsequent, large-scale trial based on definitive long-term clinical outcomes.

Because only a dose that is both safe and efficacious may be selected, the method is intrinsically superior to the conventional approach of conducting a phase I trial based on T alone followed by a phase II trial based on E alone. Because the design integrates E and T, it avoids the unnecessarily time-consuming, expensive approach of performing phase I and phase II trials separately. This is especially important in settings with low accrual rates, such as the trial of tPA for pediatric AIS described in this report. Finally, combining phase I and phase II in this way also may facilitate the often cumbersome and costly transition to a phase III trial.

Sources of Funding

Disclosures

None.

References

1. Giroud M, Lemesle M, Gouyon JB, Nivelon JL, Milan C, Dumas R. Cerebrovascular disease in children under 16 years of age in the city of Dijon, France: a study of incidence and clinical features from 1985 to 1993. *J Clin Epidemiol*. 1995;48:1343–1348.
2. Govaert P, Matthys E, Zecic A, Roelens F, Oostra A, Vanzieleghem B. Perinatal cortical infarction within middle cerebral artery trunks. *Arch Dis Child Fetal Neonatal Ed*. 2000;82:F59–F63.
3. deVeber GA, MacGregor D, Curtis R, Mayank S. Neurologic outcome in survivors of childhood arterial ischemic stroke and sinovenous thrombosis. *J Child Neurol*. 2000;15:316–324.
4. Janjua N, Nasar A, Mohammad Y, Qureshi AI. The financial burden of pediatric stroke cannot be undermined in the United States. *Stroke*. 2006;37:498.
5. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. *N Engl J Med*. 1995;333:1581–1587.
6. Intracerebral hemorrhage after intravenous t-PA therapy for ischemic stroke. The NINDS t-PA Stroke Study Group. *Stroke* 1997;28:2109–2118.
7. Khatri P, Khoury JC, Alwell K, Woo D, Kissela BM, Moomaw CJ, Miller R, Cho YJ, Kleindorfer D. Potential rtPA eligibility in children: a population-based study *Stroke*. 2006;37:641.
8. Rafay M, Pontigon A, Chiang J, Jarvis A, Silver F, MacGregor D, deVeber GA. Eligibility for hyperacute thrombolytic therapy with tissue plasminogen activator (tPA) in children. *Ann Neurol*. 2006;60:142–143.
9. Andrew M, Vegh P, Johnston M, Bowker J, Ofosu F, Mitchell L. Maturation of the hemostatic system during childhood. *Blood*. 1992;80:1998–2005.
10. Monagle P, Barnes C, Ignjatovic V, Furmedge J, Newall F, Chan A, De Rosa L, Hamilton S, Ragg P, Robinson S, Auldist A, Crock C, Roy N, Rowlands S. Developmental haemostasis: impact for clinical haemostasis laboratories. *Thromb Haemost*. 2006;95:362–372.
11. Siegbahn A, Ruusuvaara L. Age dependence of blood fibrinolytic components and the effects of low-dose oral contraceptives on coagulation and fibrinolysis in teenagers. *Thromb Haemost*. 1988;60:361–364.
12. Strolin Benedetti M, Baltes EL. Drug metabolism and disposition in children. *Fund Clin Pharmacol*. 2003;17:281–299.
13. Amlie-Lefond C, Benedict SL, Bernard T, Carpenter J, Chan A, deVeber GA, Dowling MM, Hovinga C, Ichord R, Kirkham F, Kirton A, Tsuchida T, Whelan HT, Zamel K, International Pediatric Stroke Study Investigators. Thrombolysis in children with arterial ischemic stroke: initial results from the International Pediatric Stroke Study. *Stroke*. 2007;38:485.
14. Janjua N, Nasar A, Lynch JK, Qureshi AI. Thrombolysis for ischemic stroke in children: data from the nationwide inpatient sample. *Stroke*. 2007;38:1850–1854.
15. Thall PF, Cook JD. Dose-finding based on efficacy-toxicity trade-offs. *Biometrics*. 2004;60:684–693.
16. Berry DA. Clinical trials: is the Bayesian approach ready for prime time? Yes! *Stroke*. 2005;36:1621–1622.
17. Goodman SN, Zahurak ML, Piantadosi S. Some practical improvements in the continual reassessment method for phase I studies. *Stat Med*. 1995;14:1149–1161.
18. Thall PF, Lee SJ. Practical model-based dose-finding in phase I clinical trials: methods based on toxicity. *Int J Gynecol Cancer*. 2003;13:251–261.
19. Lyden PD, Zivin JA. Hemorrhagic transformation after cerebral ischemia: mechanisms and incidence. *Cerebrovasc Brain Metab Rev*. 1993;5:1–16.
20. Wang X, Tsuji K, Lee SR, Ning M, Furie KL, Buchan AM, Lo EH. Mechanisms of hemorrhagic transformation after tissue plasminogen activator reperfusion therapy for ischemic stroke. *Stroke* 2004;35(suppl I):I-2726–I–2730.
21. Thall PF, Cook JD, Estey EH. Adaptive dose selection using efficacy-toxicity trade-offs: illustrations and practical considerations. *J Biopharm Stat.* 2006;16:623–638.
22. Thall PF. Some geometric methods for constructing decision criteria based on two-dimensional parameters. *J Stat Planning Inference*. 2008;138:516–527.
23. Rha JH, Saver JL. The impact of recanalization on ischemic stroke outcome: a meta-analysis. *Stroke*. 2007;38:967–973.
24. Tandberg Askevold E, Naess H, Thomassen L. Predictors for recanalization after intravenous thrombolysis in acute ischemic stroke. *J Stroke Cerebrovasc Dis*. 2007;16:21–24.
25. Gupta AA, Leaker M, Andrew M, Massicotte P, Liu L, Benson LN, McCrindle BW. Safety and outcomes of thrombolysis with tissue plasminogen activator for treatment of intravascular thrombosis in children. *J Pediatr*. 2001;139:682–688.
26. Gooley TA, Martin PJ, Fisher LD, Pettinger M. Simulation as a design tool for phase I/II clinical trials: an example from bone marrow transplantation. *Control ClinTrials*. 1994;15:450–462.
27. O'Quigley J, Hughes MD, Fenton T. Dose-finding designs for HIV studies. *Biometrics*. 2001;57:1018–1029.
28. Braun TM. The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes. *Control Clin Trials*. 2002;23:240–256.
29. Ivanova A. A new dose-finding design for bivariate outcomes. *Biometrics*. 2003;59:1001–1007.
30. Bekele BN, Shen Y. A Bayesian approach to jointly modeling toxicity and biomarker expression in a phase I/II dose-finding trial. *Biometrics*. 2005;61:343–354.
31. Berry DA, Mueller P, Grieve AP, Smith M, Parke T, Blazek R, Mitchard N, Krams M. Adaptive Bayesian designs for dose-ranging drug trials. *Lecture Notes Stat*. 2001;2001:99–181.

Practical Model-Based Dose Finding in Early-Phase Clinical Trials: Optimizing Tissue Plasminogen Activator Dose for Treatment of Ischemic Stroke in Children

Harry T. Whelan, John D. Cook, Catherine M. Amlie-Lefond, Collin A. Hovinga, Anthony K. Chan, Rebecca N. Ichord, Gabrielle A. deVeber and Peter F. Thall

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://stroke.ahajournals.org/content/39/9/2627