



Published in final edited form as:

*Biometrics*. 2011 December ; 67(4): 1638–1646. doi:10.1111/j.1541-0420.2011.01580.x.

## Optimizing the Concentration and Bolus of a Drug Delivered by Continuous Infusion

Peter F. Thall<sup>1,\*</sup>, Aniko Szabo<sup>2</sup>, Hoang Q. Nguyen<sup>1</sup>, Catherine M. Amlie-Lefond<sup>3</sup>, and Osama O. Zaidat<sup>3</sup>

<sup>1</sup>Department of Biostatistics, M.D. Anderson Cancer Center, Houston, TX, U.S.A.

<sup>2</sup>Department of Population Health, Medical College of Wisconsin, Milwaukee, WI, U.S.A.

<sup>3</sup>Department of Neurology, Neurosurgery and Radiology, Medical College of Wisconsin, Milwaukee, WI, U.S.A.

### Summary

We consider treatment regimes in which an agent is administered continuously at a specified concentration until either a response is achieved or a predetermined maximum infusion time is reached. Response is an event defined to characterize therapeutic efficacy. A portion of the maximum planned total amount administered is given as an initial bolus. For such regimes, the amount of the agent received by the patient depends on the time to response. An additional complication when response is evaluated periodically rather than continuously is that the response time is interval censored. We address the problem of designing a clinical trial in which such response time data and a binary indicator of toxicity are used together to jointly optimize the concentration and the size of the bolus. We propose a sequentially adaptive Bayesian design that chooses the optimal treatment for successive patients by maximizing the posterior mean utility of the joint efficacy-toxicity outcome. The methodology is illustrated by a trial in which tissue plasminogen activator is infused intra-arterially as rapid treatment for acute ischemic stroke.

### Keywords

Adaptive design; Bayesian design; Clinical trial; Continuous infusion; Phase I/II clinical trial; Stroke

## 1. Introduction

Many phase I/II designs that choose an optimal dose based on efficacy and toxicity have been proposed. Most of these methods characterize clinical outcomes as discrete variables (cf. Gooley, et al. 1994; Thall and Russell, 1998; O'Quigley, Hughes and Fenton, 2001; Braun, 2002; Ivanova, 2003; Thall and Cook, 2004; Bekele and Shen, 2005; Zhang, Sargent, and Mandrekar, 2006; Thall, Nguyen and Estey, 2008). Phase I/II methods also have been proposed based on two time-to-event outcomes (Yuan and Yin, 2009) and two ordinal outcomes (Houede, et al., 2010). While the problem that motivated the present paper is to optimize a 2-dimensional treatment based on efficacy and toxicity, the specific structure of

---

\*Corresponding author, rex@mdanderson.org .

### 7. Supplementary Materials

Supplementary Tables 1 – 14, referenced in Section 5, are available under the Paper Information link at the Biometrics website <http://www.biometrics.tibs.org>.

our setting does not fit any of the dose-finding paradigms noted above. To explain why this is the case, we first give some background on the medical setting and treatment regime.

Acute ischemic stroke (AIS) is a major cause of mortality and disability in adults (Johnson, Mendis and Mathers, 2009). A new therapeutic modality for AIS is intra-arterial (IA) fibrinolytic infusion, wherein a thrombolytic agent to dissolve the clot that caused the stroke is delivered via two telescoping catheters, one supportive in the carotid artery and a smaller microcatheter within it positioned directly into the clot. The catheters are introduced to the arterial system via a sheath placed into the femoral artery. Using live X-ray fluoroscopic guidance, the catheters are moved through the carotid artery leading to the site in the brain artery where the clot leading to AIS occurred, and the agent is infused via the microcatheter. A thrombolytic agent approved by the U.S. Food and Drug Administration for intravenous (IV) treatment of AIS in adults is tissue plasminogen activator (tPA). While effects of IV tPA in adult stroke patients are well understood, optimally safe and efficacious concentrations of IA tPA have not been established. The methodology described in this paper was motivated by the desire to design a clinical trial to optimize administration of IA tPA.

The treatment regime is as follows. For a given concentration  $c$  in mg/kg body weight and fixed maximum volume  $V$ , the maximum total dose is  $cV$ . Since  $V$  is fixed, hereafter we set  $V = 1$  without loss of generality. A proportion  $q$  of the maximum volume is given as an initial bolus at  $t = 0$ , followed by continuous infusion (ci) of the remaining proportion  $1 - q$  at a constant rate for a maximum time period  $t^*$ . In the IA tPA trial,  $t^* = 2$  hours. Efficacy is the time to dissolve the clot,  $Y_E$ . This includes the possibility that the clot is dissolved immediately by the bolus ( $Y_E = 0$ ). The ci is stopped at the time of response if it is observed by time  $t^*$ , otherwise  $Y_E$  is right-censored at  $t^*$ . Toxicity is the binary indicator  $Y_T$  of symptomatic intra-cerebral hemorrhage (SICH), characterized by neurological worsening compared to baseline in terms of a standardized stroke severity scale, and is associated with high rates of morbidity and mortality. SICH is evaluated by brain imaging, using head computerized tomography (CT) scan or magnetic resonance imaging (MRI), typically at a fixed time much later than  $t^*$ , such as 24 or 48 hours. Patients for whom the infusion fails to dissolve the clot are believed to be at higher risk of toxicity.

If response is observed continuously, then each patient's outcome data consist of  $Y_T$  and either the response time  $Y_E$  if the clot is dissolved before  $t^*$  or the fixed right-censoring time  $t^*$ . When response is evaluated periodically, which may be the only practical possibility in some settings, only the time interval in which  $Y_E$  occurred is known. In the IA tPA trial, response is evaluated at 15 minute intervals. For example, if a patient's clot was not dissolved by the 30 minute evaluation but was found to have dissolved by the 45 minute evaluation, then it is only known that the response occurred during the interval (30, 45]. Thus,  $Y_E$  is interval censored from 0 to  $t^*$  and administratively right-censored at  $t^*$ . With either continuous or periodic observation of  $Y_E$ , because the ci is stopped before  $t^*$  if the clot is dissolved, the amount of the agent that a patient actually receives depends on the patient's response time as well as  $c$  and  $q$ . Consequently, since toxicity is scored after the ci is completed the distribution of  $Y_T$  depends on  $Y_E$ .

The goal is to jointly optimize  $(c, q)$  over a design space consisting of a rectangular grid of the eight pairs obtained from the bolus proportions  $q = 0.10$  and  $0.20$ , and the concentrations  $c = 0.20, 0.30, 0.40$ , and  $0.50$  mg/kg. These  $(c, q)$  combinations were chosen by two co-authors of this paper (CMA and OZ), expert stroke, vascular, and interventional neurologists. The main problems that we address here are to (1) specify tractable probability models for  $Y_E$  and  $[Y_T|Y_E]$  as functions of  $(c, q)$  reflecting the modes of administration and observation, and (2) construct an adaptive decision procedure utilizing  $Y_T$  and the response

time data described above to choose  $(c, q)$  for successive patients. We model the marginal distribution of  $Y_E$  as a mixture of a discrete mass at  $t = 0$  for the bolus and a continuous component on  $(0, \infty)$  for the ci. The conditional toxicity probability of  $[Y_T|Y_E]$  is a nonlinear function of  $c, q$ , and  $Y_E$ . Elicited utilities of joint  $(Y_E, Y_T)$  outcomes are used as a basis for adaptively assigning treatments to successive patients as they are enrolled during the trial. Each patient is given the treatment combination  $(c, q)$  maximizing the current posterior mean utility.

The probability model is given in Section 2, followed by descriptions of the utility function and design in Section 3. Application to the IA tPA trial is described in Section 4, and simulation studies are summarized in Section 5. We close with a discussion in Section 6.

## 2. Probability Models

### 2.1 Models for Response and Toxicity

While there are many parametric regression models for event time data (cf. Ibrahim, Chen and Sinha, 2001), with this treatment regime we model efficacy as a time-to-event outcome, as follows. Since efficacy is monitored on the interval  $0 \leq t \leq t^*$ , to simplify notation we define  $Y_E$  in terms of standardized time,  $s = t/t^*$ . Denote the parameter vector of the marginal distribution of  $Y_E$  by  $\alpha$ . We define the distribution of  $[Y_E | c, q, \alpha]$  in terms of a hazard function with two components, a discrete component  $p_0(c, q, \alpha)$ , the probability that the clot is dissolved instantaneously by the bolus at  $s = 0$ , and a continuous component,  $\lambda(s, c, q, \alpha)$ , for  $s \geq 0$ . For integrated continuous hazard  $\Lambda(s, c, q, \alpha) = \int_0^s \lambda(y, c, q, \alpha) dy$ , the cumulative hazard function is  $-\log\{1 - p_0(c, q, \alpha)\} + \Lambda(s, c, q, \alpha)$  for  $s \geq 0$ . Consequently, denoting the indicator of an event  $A$  by  $\mathbf{1}(A)$ , the probability distribution function (pdf) of  $Y_E$  is the discrete-continuous mixture

$$f_E(y, c, q, \alpha) = p_0(c, q, \alpha) \mathbf{1}(y=0) + \{1 - p_0(c, q, \alpha)\} \lambda(y, c, q, \alpha) e^{-\Lambda(y, c, q, \alpha)} \mathbf{1}(y>0),$$

and the cumulative distribution function (cdf) is

$$F_E(y, c, q, \alpha) = 1 - \{1 - p_0(c, q, \alpha)\} e^{-\Lambda(y, c, q, \alpha)} \quad (1)$$

for  $y \geq 0$ . In particular,  $F_E(0, c, q, \alpha) = p_0(c, q, \alpha)$  since  $\Lambda(0, c, q, \alpha) = 0$ .

To apply this model, functional forms for  $p_0$  and  $\lambda$  must be specified. Any model used in sequential outcome-adaptive decision making based on small to moderate sized samples must balance flexibility to accurately reflect the observed data with tractability to facilitate computation. To allow  $p_0$  and  $\lambda$  to vary nonlinearly in both  $c$  and  $q$ , we will use  $c^{\alpha_1}$  and  $q^{\alpha_2}$  rather than  $c$  and  $q$  as arguments, where  $\alpha_1, \alpha_2 > 0$  are model parameters. The additional flexibility provided by  $\alpha_1$  and  $\alpha_2$  provides a basis for distinguishing more reliably among different values of  $(c, q)$  in terms of their effects on  $p_0$  and  $\lambda$  when applying the adaptive decision scheme. Based on clinical experience, we assume that

- a.  $p_0(c, q, \alpha)$  increases in both  $c$  and  $q$ , and
- b. the clot cannot dissolve instantaneously at  $s = 0$  without a bolus infusion of some tPA, hence  $p_0(c, 0, \alpha) = 0$  for all  $c > 0$  and  $p_0(0, q, \alpha) = 0$  for all  $0 \leq q \leq 1$ .

A simple, flexible function with properties (a) and (b) is

$$p_0(c, q, \alpha) = 1 - \exp(-\alpha_0 c^{\alpha_1} q^{\alpha_2}) \quad \text{for } \alpha_0 > 0. \quad (2)$$

We require the hazard function  $\lambda(s, c, q, \alpha)$  for the clot dissolving during the ci to have the following properties:

- i.  $\lambda(s, c, q, \alpha)$  must be continuous in  $s$ ;
- ii.  $\lambda(s, c, q, \alpha)$  must be sufficiently flexible so that it may be monotone increasing, monotone decreasing, or non-monotone in  $s$ ;
- iii.  $\lambda(s, 0, q, \alpha) > 0$ , to allow a non-zero baseline hazard if no tPA is given,  $c = 0$ ;
- iv.  $\lambda(s, c, 0, \alpha) > 0$  to allow the possibility that the clot is dissolved if no bolus is given;
- v. the integrated continuous hazard  $\Lambda(s, c, q, \alpha)$  must be numerically tractable;
- vi.  $\lambda(s, c, q, \alpha)$  must increase in both  $c$  and  $q$ , and may be nonlinear in either  $c$  or  $q$ .

An intuitive approach to constructing a function  $\lambda$  with these properties is to define it in terms of the *cumulative delivered dose by standardized time  $s$* , which is given by

$$d(s, c, q) = c \{q + (1 - q)s\}.$$

This function increases linearly in  $s$  with slope  $c(1 - q)$  from minimum value  $d(0, c, q) = cq$  at  $s = 0$  to  $d(1, c, q) = c$  at the last possible observation time  $s = 1$  for  $Y_E$ . While a hazard function with properties (i) – (v) may be obtained by using  $d(s, c, q)$  as an argument, to also obtain property (vi), we use the more general function

$$d(s, c^{\alpha_1}, q^{\alpha_2}) = c^{\alpha_1} \{q^{\alpha_2} + (1 - q^{\alpha_2})s\},$$

which may be considered the *effective cumulative delivered dose by standardized time  $s$* . We define the hazard function to take the form

$$\lambda(s, c, q, \alpha) = \alpha_3 + \frac{\alpha_4 \alpha_5 \{d(s, c^{\alpha_1}, q^{\alpha_2})\}^{\alpha_5 - 1}}{1 + \alpha_4 \{d(s, c^{\alpha_1}, q^{\alpha_2})\}^{\alpha_5}} \quad \text{for } s > 0, \quad (3)$$

where  $\alpha_j > 0$  for all  $j = 1, \dots, 5$ . The baseline hazard of the clot dissolving if no tPA is  $\alpha_3$ . The ratio added to  $\alpha_3$  is a log logistic hazard function with argument  $d(s, c^{\alpha_1}, q^{\alpha_2})$  and shape parameter  $\alpha_5$ . The parameter  $\alpha_1$  allows  $\lambda$  to vary nonlinearly in  $c$  while  $\alpha_2$  determines the relative magnitude of the contribution  $c^{\alpha_1}(1 - q^{\alpha_2})s$  of the ci versus the contribution  $c^{\alpha_1}q^{\alpha_2}$  of the bolus. Thus,  $\alpha = (\alpha_0, \dots, \alpha_5)$  characterizes  $p_0$  and  $\lambda$ .

Integrating (3) gives the cumulative hazard function

$$\Lambda(s, c, q, \alpha) = \alpha_3 s + \frac{1}{c^{\alpha_1} (1 - q^{\alpha_2})} \log \left[ \frac{1 + \alpha_4 \{d(s, c^{\alpha_1}, q^{\alpha_2})\}^{\alpha_5}}{1 + \alpha_4 (c^{\alpha_1} q^{\alpha_2})^{\alpha_5}} \right] \quad \text{for } s > 0. \quad (4)$$

We define the distribution of  $Y_T$  conditional on  $Y_E$  because toxicity is scored by imaging at 48 hours, after  $Y_E$  has been observed. Our model must account for the possibilities that either larger  $Y_E$ , hence a larger amount of the continuously infused agent, or failure to dissolve the clot may increase the risk of toxicity. To ensure that the probability of toxicity

is a function of  $c$ ,  $q$ , and  $Y_E$  with these properties, denoting the minimum of  $a$  and  $b$  by  $a \wedge b$  and  $\beta = (\beta_0, \dots, \beta_4)$ , we assume the model

$$\begin{aligned}\pi_T(T_E, c, q, \beta) &= \Pr(Y_T=1|Y_E, c, q, \beta) \\ &= 1 - \exp\left[-\left\{\beta_0 + \beta_2 c^{\beta_1} q + \beta_3 c^{\beta_1} (1-q)(Y_E \wedge 1) + \beta_4 \mathbf{1}(Y_E > 1)\right\}\right].\end{aligned}\quad (5)$$

Under this model,  $\beta_2 c^{\beta_1} q$  is the effect of the bolus,  $\beta_2 c^{\beta_1} (1-q)(Y_E \wedge 1)$  is the effect of the continuously infused portion,  $\beta_4$  is the effect of failing to dissolve the clot, and  $1 - e^{-\beta_0}$  is the baseline probability of toxicity if no tPA is given. The bolus size  $q$  is not exponentiated as in the definition of  $\lambda$  since this would result in an over-parameterized model for  $\pi_T$ . Thus, the model parameter vector  $\theta = (\theta, \beta)$  has dimension 11.

Although treatment begins with a bolus in the IA tPA trial, if no bolus were given then  $q = 0$  and the effective delivered dose at standardized time  $s$  would be  $d(s, c^{\alpha_1}, 0) = c^{\alpha_1} s$ . In this case,  $\alpha_2$  would be dropped from the model, the hazard function (3) would simplify to

$$\lambda(s, c, 0, \alpha) = \alpha_3 + \frac{\alpha_4 \alpha_5 (c^{\alpha_1} s)^{\alpha_5} - 1}{1 + \alpha_4 (c^{\alpha_1} s)^{\alpha_5}} \quad \text{for } s > 0,$$

and the cumulative hazard function (4) would become

$$\Lambda(s, c, 0, \alpha) = \alpha_3 s + c^{-\alpha_1} \log\left\{1 + \alpha_4 (c^{\alpha_1} s)^{\alpha_5}\right\} \quad \text{for } s > 0,$$

with  $\dim(\alpha)$  reduced from 6 to 5. The model for  $\pi_T$  would be simplified by dropping  $\beta_2$  and the linear component in (5) would be reduced to  $\beta_0 + \beta_3 c^{\beta_1} (Y_E \wedge 1) + \beta_4 \mathbf{1}(Y_E > 1)$ , with  $\dim(\beta)$  reduced from 5 to 4, so  $\beta$  would have dimension 9.

## 2.2 Joint Distribution of Response and Toxicity

Given the conditional probability  $\pi_T(y_E, c, q, \alpha)$  for  $Y_E = y_E$  and the marginal distribution  $f_E(y_E|c, q, \alpha)$  of  $Y_E$ , the joint distribution of  $\mathbf{Y} = (Y_E, Y_T)$  is given in general by

$$f_{E,T}(y_E, y_T|c, q, \theta) = f_E(y_E|c, q, \alpha) \Pr(Y_T = y_T|y_E, c, q, \beta) \quad \text{for } y_T = 0, 1 \quad \text{and } y_E \geq 0.$$

Since  $\Pr(Y_T = 1|y_E, c, q, \beta) = 1 - \Pr(Y_T = 0|y_E, c, q, \beta) = \pi_T(y_E, c, q, \alpha)$ , if  $Y_E$  is observed continuously then each patient's likelihood contribution takes the form

$$\begin{aligned}\mathcal{L}(Y)|c, q, \theta &= \left[p_0(c, q, \alpha) \pi_T(0, c, q, \beta)^{Y_T} \{1 - \pi_T(0, c, q, \beta)\}^{1-Y_T}\right]^{1(Y_E=0)} \\ &\quad \times \left[f_E(y_E|c, q, \alpha) \pi_T(y_E, c, q, \beta)^{Y_T} \{1 - \pi_T(y_E, c, q, \beta)\}^{1-Y_T}\right]^{1(0 < Y_E \leq 1)} \\ &\quad \times \left[\{1 - F_E(1|c, q, \alpha)\} \pi_T(1, c, q, \beta)^{Y_T} \{1 - \pi_T(1, c, q, \beta)\}^{1-Y_T}\right]^{1(Y_E > 1)}.\end{aligned}\quad (6)$$

The first line of expression (6) is the probability that the clot is dissolved instantaneously by the bolus, the second line is the probability that the clot is dissolved during the ci, and third line is the probability that the clot is not dissolved by standardized time  $s = 1$ , each computed either with or without toxicity, i.e.  $Y_T = 1$  or 0.

When  $Y_E$  is evaluated at the ends of successive intervals, as in the IA tPA trial, the likelihood must account for interval censoring. Given interval  $I_E = (y_E^a, y_E^b] \subset [0, 1]$ , denote

$$\pi_{E,T}(I_E, y_T | c, q, \theta) = \Pr(y_E^a < Y_E \leq y_E^b, Y_T = y_T | c, q, \theta). \quad (7)$$

This is the relevant probability if the specific value of  $Y_E$  is not observed but, instead, it is only known that the efficacy event did not occur by time  $y_E^a$  and did occur by time  $y_E^b$ . In this case, infusion is stopped at the end of the interval,  $y_E^b$ , and consequently the probability of toxicity is  $\Pi_T(y_E^b, c, q, \beta_T)$ . It follows that

$$\begin{aligned} \pi_{E,T}(I_E, y_T | c, q, \theta) &= \Pr(y_E^a < Y_E \leq y_E^b | c, q, \alpha) f_{T|E}(y_T | y_E^b, c, q, \beta) \\ &= \{F_E(y_E^b | c, q, \alpha) - F_E(y_E^a | c, q, \alpha)\} \pi_T(y_E^b, c, q, \beta)^{y_T} \{1 - \pi_T(y_E^b, c, q, \beta)\}^{1-y_T}, \end{aligned} \quad (8)$$

with  $F_E$  computed by applying formulas (1), (2), and (4) and  $\pi_T$  specified by (5).

Let  $\{I_{E,1}, \dots, I_{E,M}\}$  be a partition of  $(0, 1]$  into all possible subintervals where  $Y_E$  may be known to fall. In this case, the second term in the product that defines the likelihood (6) for continuous observation of  $Y_E$  is replaced by

$$\prod_{m=1}^M \left[ \pi_{E,T}(I_{E,m}, 1 | c, q, \theta)^{Y_T} \pi_{E,T}(I_{E,m}, 0 | c, q, \theta)^{1-Y_T} \right]^{1(Y_E \in I_{E,m})}.$$

### 3. Decision Criteria and Trial Conduct

#### 3.1 Utilities

Given one of the above likelihood formulations, tailored to the trial at hand, the problem is to construct an outcome-adaptive design for determining an optimal pair  $(c, q)$ . To do this, we take an approach similar to that of Houede, et al. (2010), who sequentially choose dose pairs in a phase I/II trial with bivariate ordinal outcomes by maximizing the posterior mean of an elicited utility. A fundamental difference that we must address here when defining utilities is that the bivariate outcome consists of a binary toxicity and continuous time to response which may be interval censored. Denote the numerical utility of outcome  $\mathbf{y}$  by  $U(\mathbf{y})$ . In practice,  $U(\mathbf{y})$  is elicited from the physicians planning the trial. We will provide details of how this was done for the IA tPA trial below. Given  $\theta$ , the mean over  $\mathbf{Y}$  of the utility for a patient who receives the treatment combination  $(c, q)$  is

$$u(c, q, \theta) = E_{\mathbf{Y}}\{U(\mathbf{Y}) | c, q, \theta\} = \sum_{y_T=0}^1 \int_{y_E=0}^{\infty} U(\mathbf{y}) f_{E,T}(\mathbf{y} | c, q, \theta) dy_E. \quad (9)$$

For each new cohort of patients during the trial, we exploit the Bayesian model by adaptively selecting the  $(c, q)$  combination that is optimal in the sense that it maximizes the posterior mean of  $u(c, q, \theta)$  based on the most recent data. Let  $\mathcal{D}_n$  denote the data from the first  $n$  patients, so that the accumulating data may be represented by the nested sequence  $\mathcal{D}_1 \subset \mathcal{D}_2 \subset \dots \subset \mathcal{D}_n \subset \dots$  as patients are treated and their outcomes are observed during

trial conduct. The optimal  $(c, q)$  maximizing the posterior mean utility given  $\mathcal{D}_n$  (Berger, 1985) is given by the equation

$$\begin{aligned}
 u(c, q)^{opt}(\mathcal{D}_n) &= \operatorname{argmax}_{c, q} E_{\theta} \{u(c, q, \theta) | \mathcal{D}_n\} \\
 &= \operatorname{argmax}_{c, q} E_{\theta} \left\{ \sum_{y_T=0}^1 \int_{y_E=0}^{\infty} U(\mathbf{y}) f_{E,T}(\mathbf{y} | c, q, \theta) dy_E | \mathcal{D}_n \right\}.
 \end{aligned} \tag{10}$$

With interval censoring due to sequential evaluation of  $Y_E$  and  $\Pr(Y_E = 0) > 0$ , given the resulting partition  $\{I_{E,0}, I_{E,1}, \dots, I_{E,M}\}$  of  $[0, 1]$  with  $I_{E,0} = \{0\}$ , a practical approach is to elicit numerical utilities for each of the  $2(M + 2)$  sets of  $\mathbf{Y}$  values obtained from the cross product  $\{I_{E,0}, I_{E,1}, \dots, I_{E,M}, I_{E,M+1}\} \times \{0, 1\}$ , where  $I_{E,M+1} = (1, \infty)$  is the outcome that the clot was not dissolved by the end of the infusion at standardized time 1. Denoting the utility of the observed outcome  $\{Y_E \in I_{E,m}, Y_T = y_T\}$  by  $U(I_{E,m}, y_T)$ , given this structure, the objective function given by (10) takes the form

$$u(c, q)^{opt}(\mathcal{D}_n) = \operatorname{argmax}_{a, q} \sum_{y_T=0}^1 \sum_{m=0}^{M+1} U(I_{E,m}, y_T) E_{\theta} \{ \pi_{E,T}(I_{E,m}, y_T | c, q, \theta) | \mathcal{D}_n \}.$$

A possible alternative to the utility-based approach might be to use a linear combination such as  $F_E(1 | c, q, \beta) - \zeta \pi_T(1 | c, q, \alpha)$  as an objective function, where  $0 < \zeta < 1$  is a design parameter quantifying the importance of achieving a response relative to suffering a toxicity. Another alternative might be a trade-off function based on  $F_E(1 | c, q, \beta)$  and  $\pi_T(1 | c, q, \alpha)$ , similar to that used by Thall and Cook (2004). These objective functions do not distinguish between responses achieved quickly or later during the infusion period, and they fail to account for effects of  $Y_E < 1$  on  $\pi_T$ . In contrast, our utility function accounts for the desirability of all observable  $(Y_E, Y_T)$  pairs. Given the data structure and goals of the IA tPA trial, choosing  $(c, q)$  to optimize the posterior mean utility is a logical and practical approach.

### 3.2 Safety and Futility Constraints

It does no good to treat patients with the  $(c, q)$  that optimizes the posterior expected utility if all pairs being considered are either excessively toxic or inefficacious. To protect patients during the trial, we impose the following safety/futility rules. Given  $\mathcal{D}_n$ , a pair  $(c, q)$  is *unacceptable* if either it is likely to be too toxic,

$$\Pr \left\{ \pi_T(1, c, q, \theta) > \bar{\pi}_T | \mathcal{D}_n \right\} > p_T, \tag{11}$$

or it is likely to be inefficacious,

$$\Pr \left\{ F_E(1, c, q, \alpha) < \bar{\pi}_E | \mathcal{D}_n \right\} > p_E, \tag{12}$$

where  $\bar{\pi}_T$  is the maximum allowed  $\pi_T(1, c, q, \theta)$  and  $\bar{\pi}_E$  is the minimum allowed  $F_E(1, c, q, \alpha)$ . The fixed limits  $\bar{\pi}_T$  and  $\bar{\pi}_E$  are elicited from the physicians, while  $p_T$  and  $p_E$  are cut-offs, usually between .80 and .99, calibrated to obtain a design with good properties. Aside from the IA tPA trial's data structure and the models underlying  $\pi_T(1, c, q, \theta)$  and  $F_E(1, c, q, \alpha)$ , the criteria (11) and (12) are similar to the phase I/II acceptability rules used by Thall and



Russell (1998), Thall and Cook (2004) and Thall, Nguyen and Estey (2008). The criterion (11) pertains to safety, and the efficacy criterion (12) is similar to Bayesian phase II futility stopping rules (Thall, Simon and Estey, 1995). Denote the set of acceptable  $(c, q)$  pairs determined by (11) and (12) based on the most recent data by  $\mathcal{A}(\mathcal{D}_n)$ . We impose the additional safety constraint that no untried concentration may be skipped when escalating, since  $\pi_T(Y_E, c, q, \beta_T)$  increases in  $c$  but may not be monotone in  $q$ . For example, if the largest value of  $c$  in  $\{0.20, 0.30, 0.40, 0.50\}$  for which patients have been treated is 0.30, then regardless of  $q$  the next cohort may be treated only at  $(c, q)$  pairs for which  $c \leq 0.40$ .

A possible alternative to using (11) and (12) might be a single criterion based on  $u(c, q, \theta)$ . One might specify a fixed lower bound  $U$  for the utility and declare a pair  $(c, q)$  to have

unacceptably low utility if  $Pr\{u(c, q, \theta) < U | \mathcal{D}_n\} > p_U$  for fixed cut-off  $p_U$ . If all  $(c, q)$  were unacceptable in this sense the trial would be stopped. While choosing a value of  $U$  might not be as intuitive as  $\bar{\Pi}_T$  and  $\bar{\Pi}_{-E}$ , since  $U(Y)$  jointly quantifies safety and efficacy the use of such a single stopping rule would provide a somewhat more unified design.

### 3.3 Trial Conduct

Once the design parameters and model are established, given a set of  $(c, q)$  pairs, maximum sample size,  $N$ , and cohort size, the trial is conducted as follows. The first cohort is treated at a starting  $(c, q)$  combination chosen by the physicians, and the choice may be guided by the numerical utilities and prior means. While the usual fear in phase I where only toxicity is considered is overdosing the first few patients, in the present setting when choosing the starting  $(c, q)$  pair this fear may be counterbalanced by the concern that patients may be given too little tPA to dissolve their clots. A given  $c$  may be too low to dissolve the clot that caused the stroke but high enough to cause a variety of adverse effects not associated with observable hemorrhage (SICH) and not easily be detected, and thus such events cannot feasibly be included in an outcome-adaptive decision making procedure. Such adverse effects include cytotoxicity, degradation of extracellular matrix, and increased permeability of the neurovascular unit with the development of cerebral edema (Kaur, et al., 2004). For each cohort after the first, if  $\mathcal{A}(\mathcal{D}_n)$  is empty then the trial is stopped early with the conclusion that all  $(c, q)$  pairs are unacceptable. If  $\mathcal{A}(\mathcal{D}_n)$  is not empty then the next cohort is treated at the best acceptable pair  $(c, q)^{opt}(\mathcal{D}_n)$ , subject to the do-not-skip rule. At the end of the trial, the  $(c, q)^{opt}(data)$  pair based on the final data is selected.

## 4. Application to the IA tPA Trial

### 4.1 Utilities and Design Parameters

To evaluate  $Y_E$ , the clot is imaged at the start of infusion when the bolus is given, and thereafter every 15 minutes up to the maximum infusion duration of  $t^* = 120$  minutes. The planned observation intervals are thus  $[0, 15], \dots, (105, 120]$  and  $(120, \infty]$ . These are given in Table 1 along with the corresponding intervals in standardized time. Since  $Y_E$  is observed only at the end of each observation interval up to the last, as a covariate in the linear term of  $\pi_T$  it can take on only the values of the interval endpoints, also given in Table 1, unless it appears in the indicator  $1(Y_E > 1)$ . If the evaluation times deviate from this planned schedule for some patients, however, then the likelihood can easily be modified to accommodate this, so that the actual interval observation data on  $Y_E$  will be recorded during the trial.

Up to  $N = 36$  patients will be treated in cohorts of size 1, with the aim to choose the  $(c, q)$  pair that maximizes the posterior mean utility among the set of eight possible pairs obtained from  $q = 0.10, 0.20$  and  $c = 0.20, 0.30, 0.40, 0.50$ . The maximum sample size was chosen based on an anticipated accrual rate of 1 patient/month/site with 15 sites participating and



5% of accrued patients both eligible and consenting. This would give .75 patients per month, so a 36-patient trial would last 48 months. The admissibility limits were specified to be  $\bar{\pi}_T = 0.15$  and  $\bar{\pi}_E = 0.50$ , and the probability cut-offs used in (11) and (12) were  $p_E = p_T = .95$ . The starting pair is  $(c, q) = (0.20, 0.10)$ , and  $N$  was chosen based on anticipated accrual limitations and the desire to complete the trial within a reasonable timeframe.

Numerical utilities, on a scale of 0 to 100, elicited for each combination of  $Y_T = 0$  or 1 and observation interval for  $Y_E$ , are given in Table 1. CMA and OZ provided the numerical utilities, acceptability limits,  $\bar{\pi}_T$ , and  $\bar{\pi}_E$ , maximum sample size,  $N$ , and means of the efficacy and toxicity events used to establish the prior, described below in Section 4.2. The elicited utilities are based on clinical experience and published data on IA therapy for AIS. The elicitation was carried out in two stages, with refinement in the second stage to ensure that the utilities decreased strictly with the time interval required to dissolve the clot. In a trial for interventional management of AIS, Khatri, et al. (2009) showed that the chance of good clinical outcome decreased greatly with longer time from symptom onset to initiation of treatment. In Table 1, the large drop in utilities when SICH occurs quantifies the fact that SICH is associated with much worse clinical outcome, as close to 50% of all SICHs are fatal.

#### 4.2 Establishing Priors

To establish priors, we used a three-step strategy. First, we elicited a large number of prior means of the probabilities  $p_0(c, q, \theta)$ ,  $F_E(s, c, q, \theta)$ , and  $\pi_T(y_E/c, q, \beta)$ . Second, based on the elicited values we repeatedly simulated a large pseudo data set and, starting with a non-informative pseudo prior, used the average of the means of the resulting pseudo posteriors as the prior means. Third, we calibrated the prior variances of the entries  $\theta$  of to obtain a prior with desirable properties.

In step 1, for each of the eight  $(c, q)$  combinations, we elicited the prior means of the probability of dissolving the clot immediately with the bolus,  $p_0(c, q, \theta)$ , within 60 minutes,

$F_E\left(\frac{1}{2} | c, q, \theta\right)$ , or within 120 minutes,  $F_E(1 | c, q, \theta)$ . Similarly, we elicited the probability of toxicity,  $\pi_T(y_E, c, q, \theta)$ , if the clot was dissolved instantaneously ( $y_E = 0$ ), if it was dissolved within the 120 minute infusion ( $y_E = 1$ ), or if it was not dissolved during the infusion ( $y_E > 1$ ). The elicited values are given in Table 2a. The last line of Table 2a expresses the prior expectation that, for  $c = .50$  and either  $q = .10$  or  $.20$ , the mean of  $\pi_T$  will increase from  $.12$  (the value for  $y_E = 1$ ) to  $.15$  if the clot is not dissolved by the end of the infusion.

For the second step, the elicited values were treated like the true state of nature and used to simulate 1000 large pseudo samples, each of size 400 with exactly 50 patients for each  $(c, q)$  combination. Starting with a very non-informative pseudo-prior on  $\theta$  in which the logarithm of each entry followed a normal distribution with mean 0 and standard deviation 20, we used the pseudo data set to compute a pseudo posterior. The average of the 1000 pseudo posterior means were used as the prior means. These are given in Table 2b. The pseudo sample size of 400 was chosen to be large enough so that prior means obtained in this way would not change substantively with a larger pseudo sample size.

For the third step, we calibrated the variances of the entries of  $\theta$  to ensure a prior that was suitably non-informative in terms of the prior effective sample sizes (ESSs) of  $\pi_T(s, c, q, \theta)$  and  $F_E(s, c, q, \theta)$ , and also to obtain good simulated performance of the design in the actual trial across a diverse set of scenarios. Denoting  $\pi_T(s, c, q, \theta)$  or  $F_E(s, c, q, \theta)$  for  $s = 0$  or 1 by  $p(\theta)$ , the ESS of the prior on  $p(\theta)$  was approximated by matching its mean and variance

with those of a beta( $a, b$ ) distribution with mean  $a/(a+b)$  and variance  $ab/\{(a+b)^2(a+b+1)\}$  and approximating the ESS as  $a+b \pi \approx E\{p(\boldsymbol{\theta})\}[1 - E\{p(\boldsymbol{\theta})\}]/\text{var}\{p(\boldsymbol{\theta})\} - 1$ . On this basis, we set the variance  $\sigma^2$  of  $\log(\theta_j)$  for each element  $\theta_j$  of  $\boldsymbol{\theta}$  to equal  $\sigma^2 = 81$ , which gave ESS values of each  $p(\boldsymbol{\theta})$  ranging from 0.17 to 0.22 with mean 0.19. It is inappropriate to specify arbitrarily large  $\sigma^2$ , since this may produce priors on  $F_E(y, c, q, \boldsymbol{\alpha})$  and  $\pi_T(y_E/c, q, \boldsymbol{\beta})$  with very large probability masses below .01 and above .99. For such a prior, the design behaves pathologically and makes erroneous decisions for patients enrolled early in the trial.

### 4.3 Posterior Computation

The computations for each interim decision include obtaining the posterior probabilities in the admissibility criteria (11) and (12) and posterior mean utility (9) for all  $(c, q)$  combinations. We used Markov chain Monte Carlo (MCMC) with Gibbs sampling (Robert and Casella, 1999) to compute all posterior quantities, based on the full conditionals. Each series of sample parameters  $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(N)}$  distributed proportionally to the posterior integrand was initialized at the mode using the two-level algorithm described in Braun et al. (2007). Because each sample chain was initialized at the mode, which reliably identifies the region of highest posterior probability density, we did not require any burn-in, and a single chain was used for each posterior computation. We used MCMC sample size  $N = 2,000$  for the dose assignments during the trial, and  $N = 16,000$  for the dose selection at the end of the trial. For each sample  $\boldsymbol{\theta}^{(i)} = (\boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)})$ , we computed  $p_0(c, q, \boldsymbol{\alpha}^{(i)})$ , then  $\Lambda(Y_E, c, q, \boldsymbol{\alpha}^{(i)})$ ,  $F_E(Y_E, c, q, \boldsymbol{\alpha}^{(i)})$  and  $\pi_T(Y_E, c, q, \boldsymbol{\beta}^{(i)})$  at every interval endpoint  $Y_E$  and for every  $(c, q)$ , and computed  $\pi_{E,T}(I_{E,m}, y_T | c, q, \boldsymbol{\theta}^{(i)})$  from (8) for each  $I_{E,m}$  and  $y_T \in \{0, 1\}$ . The elicited utilities were averaged over each of these distributions to obtain a utility for every  $(c, q)$  given  $\boldsymbol{\theta}^{(i)}$ . We computed the  $(c, q)$  posterior mean utilities by averaging over the samples  $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(N)}$ . The Monte Carlo standard error (MCSE) was computed using the batch-means method for  $F_E(1, c, q)$ ,  $\pi_T(1, c, q)$  and  $u(c, q)$  at the highest and lowest  $(c, q)$  combinations. The ratios of the MCSE to the posterior standard deviation of these quantities were small ( $< 3\%$ ), indicating MCMC convergence.

## 5. Simulation Studies

Each trial was simulated 10,000 times under each of a wide variety of scenarios. Since each scenario was specified in terms of fixed values of the marginal probabilities  $\pi_T(s, c, q)^{true}$  and  $F_E(s, c, q)^{true}$  for  $s = 0$  and 1, to obtain fixed true probabilities for all  $s$  in  $[0, 1]$  we used several interpolation methods, allowing  $\pi_T(s, c, q)^{true}$  and  $F_E(s, c, q)^{true}$  to take various shapes as functions of  $s$ . In terms of  $\pi_T(s, c, q)^{true}$ , the intermediate probabilities were interpolated as  $\pi_T(s, c, q)^{true} = \pi_T(0, c, q) + \{\pi_T(1, c, q) - \pi_T(0, c, q)\}s^\phi$  for various  $\phi$ , with  $\phi = 1$  giving linear interpolation,  $\phi = 2$  giving values “below linear”, and  $\phi = 0.5$  giving values “above linear”. We also used a method that sets  $\pi_T(0.5, c, q)^{true} = 0.5\{\pi_T(0, c, q) + \pi_T(1, c, q)\}$  and interpolates using  $\pi_T(s, c, q)^{true} = \pi_T(0, c, q) + \{\pi_T(0.5, c, q)^{true} - \pi_T(0, c, q)\}(2s)^2$  for  $0 \leq s \leq .5$  and  $\pi_T(.5, c, q) + \{\pi_T(1, c, q)^{true} - \pi_T(.5, c, q)\}(2s - 1)^2$  for  $.5 \leq s \leq 1$ , to give an “S-shaped” curve. Interpolated values of  $F_E(s, c, q)^{true}$  were obtained similarly. The joint probabilities  $\pi_{E,T}(s, c, q)^{true}$  used to generate  $(Y_E, Y_T)$  given the assigned  $(c, q)$  were computed using (7) and (8), and the resulting true utility  $u(c, q)^{true}$  was obtained from expression (9).

The simulation scenarios are given in Supplementary Tables 1 – 6. Scenario 1 uses the elicited prior means, with the utilities increasing with both  $c$  and  $q$ , so that  $(c, q) = (0.5, 0.2)$  is optimal. Scenario 2 has a similar pattern, but with a larger increase as  $(c, q)$  goes from  $(0.2, 0.1)$  to  $(0.5, 0.2)$ . In Scenario 3, the middle values  $c = 0.3$  and  $0.4$  have the highest utilities, also with  $u(c, 0.1)^{true} > u(c, 0.2)^{true}$  so a smaller bolus is more desirable. In Scenario 4, smaller values of both  $c$  and  $q$  have higher  $u(c, q)^{true}$ . Scenario 5 is unsafe, with

unacceptably high values of all  $\pi_T(s, c, q)^{true}$  compared to the upper limit  $\bar{\pi}_T = 0.15$ . In Scenario 6, all efficacy probabilities  $F_E(s, c, q)^{true}$  of dissolving the clot are unacceptably small compared to the lower limit  $\underline{\pi}_E = 0.50$ . Thus, in the last two scenarios, it is most desirable to stop the trial early and select no  $(c, q)$  pair.

To summarize the method's overall behavior, we used the following statistic. For each scenario, let  $u^{true,sel}$  denote the true utility of the final selected  $(c, q)$  and  $u_{max}$  and  $u_{min}$  the largest and smallest true utilities among all  $(c, q)$  pairs. The summary statistic  $R = (u^{true,sel} - u_{min}) / (u_{max} - u_{min})$  is the proportion of the difference between the best and worst possible choices achieved by the selected treatment. Thus,  $0 \leq R \leq 1$ , with larger values corresponding to better performance of the method. In cases such as Scenarios 5 and 6 where no treatment is acceptable, the value of  $R$  is not relevant since it is not useful to choose a treatment maximizing the utility if all treatments are unacceptable. The simulations are summarized in Table 3. Under Scenarios 1 – 4, searching among the eight possible  $(c, q)$  pairs, the method does a reliable job of selecting pairs with higher utilities, and subsample sizes are favorably balanced toward more desirable pairs. The results for Scenarios 5 and 6 show that if no pair is acceptably safe and efficacious the method is likely to stop early and select no pair. In Scenarios 1 – 4 where an acceptable  $(c, q)$  pair exists,  $R$  increases with  $N$ . The values of “% none” selected increase with  $N$ . These are fixed at 2% for all  $N$  studied under Scenario 2, and reach a maximum of 100% by  $N = 240$  in the Scenarios 5 and 6 where no pairs are acceptable.

We assessed sensitivity to the prior,  $N$ , cohort size, and  $\sigma$ , summarized in Supplementary Tables 7, 8, and 9. Supplementary Table 7 shows that, for  $N = 24$  to 240, using a prior with mean 0 and  $\sigma^2 = 20$  for all  $\log(\theta_j)$  substantially degrades  $R$  under Scenarios 1 and 2, increases  $R$  under Scenarios 3 and 4, and increases the futility stopping probabilities under Scenario 6. For  $N = 36$ , there is no general pattern of  $R$  or stopping probability with cohort size 1, 2, 3, or with  $\sigma = 7$  to 20. Sensitivity to the four interpolation methods for obtaining each scenario's probabilities between successive evaluation times  $s = 0, 0.125, \dots, 1.0$  is summarized in Supplementary Table 10. The stopping probabilities are insensitive to the interpolation method, but  $R$  may change very little or substantively, depending on the scenario and method. This is because each interpolation method gives different shapes of  $\pi_T(s, c, q)^{true}$  and  $F_E(s, c, q)^{true}$  and thus a different version of each scenario, which in turn changes  $u(c, q, \theta)^{true}$  and thus the method's behavior and  $R$ . Supplementary Table 11 summarizes sensitivity to  $\beta_4^{true}$ , the effect on  $\pi_T$  of failure to dissolve the clot, for  $\beta_4^{true} = kE_{prior}(\beta_4)$ , with  $k = 1$  to 9. In all scenarios, “% none” selected increases with  $k$ , a very desirable property. To assess the design's performance under a more parsimonious model, we simplified the mixture distribution given by (1) and (2) by fixing  $\alpha_1 = \alpha_2 \equiv \beta_1 \equiv 1$  and  $\beta_0 \equiv \beta_4 = 0$ , so that  $\pi_T(y_E) = 1 - e^{-\beta_2 c q - \beta_3 c(1-q)(y_E; 1)}$  and  $p_0 = 1 - e^{\alpha_0 c q}$ , and assumed the Weibull hazard  $\lambda(y, c, q, \alpha) = \{\alpha_3 c q + \alpha_4 c(1 - q)\} \psi y^{\psi - 1}$ , in place of (5). This model has 6 parameters,  $(\alpha_0, \alpha_4, \alpha_3, \psi, \beta_2, \beta_3)$ , compared to 11 for the original model. The simulations, in Supplementary Table 12, show that using this simpler model greatly degrades the method's performance in Scenarios 3 and 4 where the highest concentration is not the best choice, and reduces the reliability of the safety and futility rules, with early stopping probability reduced from .91 to .69 under scenario 5 and from .83 to .70 under Scenario 6.

Table 4 gives a hypothetical case-by-case example to illustrate how a trial might play out in practice, under a scenario with the best  $c$  in the middle. For each patient, the treatment values and outcomes are given with the posterior mean utilities of the eight  $(c, q)$  combinations. To conserve space, results are given for the first 12 patients and thereafter each sixth patient. The values for all 36 patients are given in Supplementary Table 13. For this example, posterior means and standard deviations of the elements of  $\theta$ , and of  $F_E(1, c,$

$q, \theta$ ) and  $\pi_T(1, c, q, \theta)$ , with prior values for comparison, are given in Supplementary Table 14.

## 6. Discussion

The methodology proposed here may be extended to oncology settings. For example, suitable modifications of the methodology may be used for a chemotherapeutic anti-cancer agent administered by ci, with a possible initial bolus, with the tumor imaged repeatedly and therapy stopped when tumor response is achieved. In such settings, the time frame likely would be much longer than the 120 minute schedule considered here, and the infusion typically would include successive cycles with interim rest periods. Additionally, toxicity might be a time-to-event variable, possibly occurring during infusion and causing treatment to be suspended or permanently stopped. Such differences are non-trivial, however, and would require substantive modifications of  $\lambda$ ,  $\pi_T$ , and the decision rules.

A simpler version of the method currently is being applied to plan a similar trial of IA tPA in pediatric stroke patients. Although pediatric AIS is rare, over 75% of children with acute AIS will die or suffer long-term neurological deficits (deVeber, et al., 2000). In this trial, it was decided to fix  $q \equiv 0.10$  since a bolus of size  $q = 0.20$  or larger was considered too risky for children. For the model, the response hazard is simplified by fixing  $\alpha_2 \equiv 1$ . The design space consists of the four concentrations  $\{0.20, 0.30, 0.40, 0.50\}$  and  $c$  is chosen based on  $u(c^{opt})(\mathcal{D}_n)$ , defined as the maximum over  $c$  of  $E_\theta\{u(c, \theta) | \mathcal{D}_n\}$ .

A computer program, named “CiBolus,” to implement this methodology is available from the website <https://biostatistics.mdanderson.org/SoftwareDownload>.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors thank the editor, associate editor, and a referee for their detailed and constructive comments that substantively improved the manuscript. The research of PT and HN was supported by NCI grant RO1-CA-83932.

## References

- Bekele BN, Shen Y. A Bayesian approach to jointly modeling toxicity and biomarker expression in a phase I/II dose-finding trial. *Biometrics*. 2005; 60:343–354. [PubMed: 16011680]
- Berger, James O. *Statistical Decision Theory and Bayesian Analysis*. 2nd Edition. Springer-Verlag; New York: 1985.
- Braun TM. The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes. *Contemporary Clinical Trials*. 2002; 23:240–256.
- Braun TM, Thall PF, Nguyen H, de Lima M. Simultaneously optimizing dose and schedule of a new cytotoxic agent. *Clinical Trials*. 2007; 4:113–124. [PubMed: 17456511]
- deVeber G, MacGregor D, Curtis R, Mayank S. Neurologic outcome in survivors of childhood arterial ischemic stroke and sinovenous thrombosis. *Journal of Childhood Neurology*. 2000; 15(5):316–324.
- 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. *Controlled Clinical Trials*. 1994; 15:450–462. [PubMed: 7851107]
- Houede N, Thall PF, Nguyen H, Paoletti X, Kramar A. Utility-based optimization of combination therapy using ordinal toxicity and efficacy in phase I/II trials. *Biometrics*. 2010; 66:532–540. [PubMed: 19673865]

- Ibrahim, JG.; Chen, M-H.; Sinha, D. Bayesian Survival Analysis. Springer; New York: 2001.
- Ivanova A. A new dose-finding design for bivariate outcomes. *Biometrics*. 2003; 59:1001–1007. [PubMed: 14969479]
- Johnson SC, Mendis S, Mathers CD. Global variation in stroke burden and mortality: estimates from monitoring, surveillance, and modelling. *The Lancet Neurology*. 2009; 8:345–354.
- Kaur J, Zhao Z, Klein GM, Lo EH, Buchan AM. The neurotoxicity of tissue plasminogen activator? *J. Cerebral Blood Flow Metabolism*. 2004; 24:945–963.
- Khatri P, Abruzzo T, Yeatts SD, Nichols C, Broderick JP, Tomsick TA, IMS I and II Investigators. Good clinical outcome after ischemic stroke with successful revascularization is time-dependent. *Neurology*. 2009; 73:1066–1072. [PubMed: 19786699]
- O’Quigley J, Hughes MD, Fenton T. Dose-finding designs for HIV studies. *Biometrics*. 2001; 57:1018–1029. [PubMed: 11764240]
- Robert, CP.; Cassella, G. Monte Carlo Statistical Methods. Springer; New York: 1999.
- Thall PF, Cook JD. Dose-finding based on efficacy-toxicity trade-offs. *Biometrics*. 2004; 60:684–693. [PubMed: 15339291]
- Thall PF, Nguyen H, Estey EH. Patient-specific dose-finding based on bivariate outcomes and covariates. *Biometrics*. 2008; 64:1126–1136. [PubMed: 18355387]
- Thall PF, Russell KT. A strategy for dose finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Biometrics*. 1998; 54:251–264. [PubMed: 9544520]
- Thall PF, Simon R, Estey EH. Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. *Stat in Medicine*. 1995; 14:357–379.
- Yuan Y, Yin G. Bayesian dose-finding by jointly modeling toxicity and efficacy as time-to-event outcomes. *Journal of the Royal Statistical Society, Series C*. 2009; 58:954–968.
- Zhang W, Sargent DJ, Mandrekar S. An adaptive dose-finding design incorporating both efficacy and toxicity. *Statistics In Medicine*. 2006; 25:2365–2383. [PubMed: 16220478]

Elicited utilities of the joint outcomes and values of  $Y_E$  used to compute  $\pi_T(Y_E, c, q, \beta)$  when efficacy is monitored in 15-minute intervals. Efficacy is defined as the time to blood clot being dissolved, and the adverse event is symptomatic intra-cerebral hemorrhage (SICH), evaluated at 48 hours. The category “> 120” corresponds to the clot not being dissolved by the end of the 120 minute infusion

**Table 1**

	<i>Time Required to Dissolve the Blood Clot</i>										
Minutes	0	1-15	16-30	31-45	46-60	61-75	76-90	91-105	106-120	> 120	
No SICH	100	95	90	85	80	75	70	60	50	30	
SICH	7	6.5	6	5	4.5	4	2	1	0	0	
		<i>Value of <math>Y_E</math> Used to Compute <math>\pi_T(Y_E, c, q, \beta)</math></i>									
	0	.125	.250	.375	.500	.625	.750	.875	1.00	1.00	1.00



**Table 2**

Elicited prior mean probabilities for each  $(c, q)$  combination studied in the IA tPA trial, and resulting parameter prior means

<u>a. Elicited Prior Mean Probabilities</u>		$c = 0.20$	$c = 0.30$	$c = 0.40$	$c = 0.50$
$q = .10$	$E\{p_0(c, q, \theta)\}$	0.10	0.15	0.15	0.25
	$E\{F_E^{1/2}   c, q, \theta\}$	0.25	0.30	0.45	0.50
	$E\{F_E   c, q, \theta\}$	0.35	0.45	0.60	0.70
$q = .20$	$E\{\pi_T(0, c, q, \theta)\}$	0.02	0.03	0.03	0.03
	$E\{\pi_T(1, c, q, \theta)\}$	0.04	0.06	0.08	0.12
	$E\{p_0(c, q, \theta)\}$	0.15	0.20	0.25	0.30
	$E\{F_E^{1/2}   c, q, \theta\}$	0.40	0.45	0.50	0.60
	$E\{F_E   c, q, \theta\}$	0.50	0.60	0.70	0.80
$q = .10 \text{ or } .20$	$E\{\pi_T(0, c, q, \theta)\}$	0.02	0.03	0.03	0.03
	$E\{\pi_T(1, c, q, \theta)\}$	0.04	0.06	0.08	0.12
	$E\{\pi_T(.50, q, \mathbf{1}(Y_E > 1), \theta)\} = .15$				

  

<u>b. Prior Mean of the Log of Each Model Parameter</u>										
$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
-1.04	-1.60	-7.25	-4.74	-2.85	2.37	-6.10	-3.79	-7.05	-5.42	-7.88

Simulation results for the IA tPA trial. Under each scenario,  $u^{true}(c, q)$  denotes the expected utility of treating a patient with the combination  $(c, q)$ . The value of the utility for the combination with the highest utility is highlighted in bold. Utilities of combinations that are too toxic or have insufficient effectiveness in the scenario have a gray background

Table 3

Scenario	q	c					% none
		0.2	0.3	0.4	0.5		
1 Prior means	0.1	$u^{true}(c, q)$ % Sel (No. Pats.)	46.9 2% (3.2)	51.5 2% (1.6)	59.2 7% (3.4)	64.4 29% (8.7)	6
	0.2	$u^{true}(c, q)$ % Sel (No. Pats.)	56.1 2% (0.8)	60.5 2% (0.8)	65.1 11% (4.5)	<b>70.6</b> 38% (11.5)	
2 Safe, high c and q=2 best	0.1	$u^{true}(c, q)$ % Sel (No. Pats.)	49.0 1% (2.6)	54.9 1% (1.2)	62.4 4% (2.3)	71.5 17% (5.9)	2
	0.2	$u^{true}(c, q)$ % Sel (No. Pats.)	52.6 1% (0.5)	58.4 1% (0.6)	65.9 13% (5.3)	<b>73.8</b> 60% (16.9)	
3 Safe, middle c and q=1 best	0.1	$u^{true}(c, q)$ % Sel (No. Pats.)	57.3 3% (2.4)	68.9 8% (3.6)	70.1 33% (9.0)	64.0 33% (9.3)	9
	0.2	$u^{true}(c, q)$ % Sel (No. Pats.)	57.1 2% (0.5)	68.6 2% (0.9)	69.7 7% (5.4)	63.5 3% (2.8)	
4 Safe, low c and q=1 best	0.1	$u^{true}(c, q)$ % Sel (No. Pats.)	<b>61.1</b> 43% (14.4)	58.7 7% (3.3)	51.6 6% (3.4)	48.0 5% (2.8)	12
	0.2	$u^{true}(c, q)$ % Sel (No. Pats.)	58.2 22% (5.0)	53.9 3% (1.5)	49.5 2% (2.2)	45.0 1% (0.9)	
5 Unsafe	0.1	$u^{true}(c, q)$ % Sel (No. Pats.)	44.8 4% (6.5)	45.2 1% (1.5)	<b>45.2</b> 1% (2.0)	45.0 1% (2.2)	91
	0.2	$u^{true}(c, q)$ % Sel (No. Pats.)	45.2 1% (0.9)	45.2 0% (0.4)	45.0 0% (0.8)	44.3 0% (0.6)	
6 Safe, but no (c, q) acceptable	0.1	$u^{true}(c, q)$ % Sel (No. Pats.)	38.2 0% (2.8)	40.0 0% (1.1)	41.9 1% (1.6)	43.3 7% (5.2)	83
	0.2	$u^{true}(c, q)$ % Sel (No. Pats.)	39.3 0% (0.4)	41.2 0% (0.4)	43.1 1% (1.4)	<b>44.4</b> 7% (4.3)	

**Table 4**

Hypothetical case-by-case example of a 36 patient trial. Data are given for patients 1 – 12, and for each sixth patient thereafter

Patient	Treatment		Outcomes		Posterior Mean Utility for Each (c, q)									
	c	q	$Y_E$	$Y_T$	(.2,.1)	(.3,.1)	(.4,.1)	(.5,.1)	(.2,.2)	(.3,.2)	(.4,.2)	(.5,.2)		
(Prior)	-	-	-	-	42.4	42.4	42.4	42.4	42.6	42.0	42.0	41.9	42.0	
1	0.2	0.1	45	No	67.7	66.6	66.3	64.1	67.3	66.2	65.9	63.6		
2	0.2	0.1	45	No	75.6	73.2	71.1	66.6	74.6	72.6	70.6	66.3		
3	0.2	0.1	No	No	59.6	59.0	59.3	58.17	59.4	58.8	59.0	57.8		
4	0.2	0.1	No	No	52.3	52.2	52.7	52.8	52.1	52.0	52.5	52.5		
5	0.2	0.1	No	No	47.8	47.9	48.7	49.2	47.7	47.8	48.6	49.0		
6	0.3	0.1	75	No	52.3	52.8	53.3	53.1	52.2	52.7	53.2	52.9		
7	0.4	0.1	90	No	54.8	55.2	55.6	55.3	54.8	55.2	55.5	55.1		
8	0.4	0.1	105	No	56.2	56.6	57.1	56.6	56.2	56.7	57.2	56.5		
9	0.4	0.2	0	No	59.4	60.3	62.5	63.5	60.9	61.9	64.3	65.1		
10	0.5	0.2	90	Yes	60.1	60.8	61.1	26.9	61.0	61.7	61.9	26.7		
11	0.4	0.2	120	No	59.9	60.6	61.3	26.0	61.0	61.8	62.4	26.1		
12	0.4	0.2	No	Yes	54.8	54.8	53.0	33.2	54.8	54.8	52.5	32.2		
18	0.3	0.1	120	No	61.3	61.8	59.6	35.0	61.6	61.9	59.1	33.7		
24	0.3	0.1	No	No	57.6	57.9	57.0	39.4	57.3	57.5	56.1	37.6		
30	0.3	0.1	105	No	56.4	56.8	56.3	37.5	56.2	56.5	55.8	36.0		
36	0.3	0.2	45	No	58.8	59.5	59.2	32.4	59.4	60.1	59.6	31.4		