Published in final edited form as:

J Stat Plan Inference. 2008 ; 138(2): 516–527. doi:10.1016/j.jspi.2007.06.013.

Some Geometric Methods for Constructing Decision Criteria Based On Two-Dimensional Parameters

Peter F. Thall

Department of Biostatistics, University of Texas, M.D. Anderson Cancer Center, 1515 Holcombe
Boulevard, Houston, Texas 77030, USA, email: rex@mdanderson.org

Abstract

This paper reviews two types of geometric methods proposed in recent years for defining statistical decision rules based on 2-dimensional parameters that characterize treatment effect in a medical setting. A common example is that of making decisions, such as comparing treatments or selecting a best dose, based on both the probability of efficacy and the probability toxicity. In most applications, the 2-dimensional parameter is defined in terms of a model parameter of higher dimension including effects of treatment and possibly covariates. Each method uses a geometric construct in the 2-dimensional parameter space based on a set of elicited parameter pairs as a basis for defining decision rules. The first construct is a family of contours that partitions the parameter space, with the contours constructed so that all parameter pairs on a given contour are equally desirable. The partition is used to define statistical decision rules that discriminate between parameter pairs in term of their desirabilities. The second construct is a convex 2-dimensional set of desirable parameter pairs, with decisions based on posterior probabilities of this set for given combinations of treatments and covariates under a Bayesian formulation. A general framework for all of these methods is provided, and each method is illustrated by one or more applications.

Keywords

Bayesian statistics; Clinical trials; Dose-finding; Indifference set; Medical decision making; Phase II clinical trial; Trade-offs

1. Introduction

Most medical treatments have multiple effects on patients. Typically, a drug's desirable anti-disease effect is called "response" or "efficacy," while undesirable effects are called "toxicity" or "adverse events." When choosing among established treatments for an individual patient, or when evaluating treatments in a clinical trial, it is useful to consider the rates of such effects jointly. The widespread use of aspirin for its analgesic effects and its ability to lower the risk of cardiovascular disease is based on the idea that the likelihoods of these beneficial effects are a desirable trade-off for the risk that aspirin also may cause an upset stomach. Beta blockers effectively reduce abnormally high blood pressure but also may cause dizziness, fatigue, insomnia or a reduced libido. How one classifies outcomes as being desirable or undesirable depends on one's treatment goals. For example, a particular drug may be characterized either as an antihistamine that causes drowsiness, or as a sleeping

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

pill that dries up one's sinuses. Consideration of trade-offs between desirable and undesirable clinical outcomes is especially important in oncology, where the goal is ideally to cure or hopefully to slow the growth of the patient's cancer, but most treatments also carry a substantial risk of harmful effects. For example, treating a hematologic malignancy with transplanted allogeneic blood or marrow stem cells obtained from a human leukocyte antigen (HLA) matched donor may have the desirable effect of killing cancer cells but also may kill normal cells in the skin, liver or gastro-intestinal tract through an autoimmune reaction known as graft-versus-host disease. Rapid treatment of stroke with agents aimed at dissolving the blood clot that caused the stroke also carries the risk that the agent may cause bleeding in the brain. A general problem arising in treatment of life-threatening diseases is that some therapies may extend average survival time but also may reduce the patient's quality of life.

This paper will focus primarily on settings where it is sensible to characterize treatment effect as a 2-dimensional parameter. The following simple example will serve to motivate the general problem and illustrate some of the main ideas. Consider a clinical setting where the observed outcome includes the pair $\mathbf{Y} = (Y_1, Y_2)$ with Y_1 the indicator of response to treatment, "efficacy," and Y_2 the indicator of a treatment-related severe adverse event (SAE), "toxicity." Denoting $\theta_{a,b} = \Pr(Y_1 = a, Y_2 = b)$ for $a, b \in \{0, 1\}$, a reasonable likelihood for $\mathbf{Y}$ is the product multinomial

$$f(\mathbf{Y}|\boldsymbol{\theta}) = \theta_{1,1}^{Y_1 Y_2} \theta_{1,0}^{Y_1(1-Y_2)} \theta_{0,1}^{(1-Y_1)Y_2} \theta_{0,0}^{(1-Y_1)(1-Y_2)}$$

with 3-dimensional parameter vector $\boldsymbol{\theta} = (\theta_{1,1}, \theta_{1,0}, \theta_{0,1})$, writing $\theta_{0,0} = 1 - \theta_{1,1} - \theta_{1,0} - \theta_{0,1}$. In this setting, before constructing statistical decision rules it often is appropriate to first apply the mapping

$$\boldsymbol{\theta} \rightarrow \boldsymbol{\xi} = (\theta_{1,1} + \theta_{1,0}, \theta_{1,1} + \theta_{0,1})$$

to reduce $\boldsymbol{\theta}$ to the 2-dimensional parameter $\boldsymbol{\xi} = (\Pr(Y_1=1), \Pr(Y_2=1))$ consisting of the marginal probabilities of response and toxicity. Thus, $\xi_1 = \Pr(\text{efficacy})$ and $\xi_2 = \Pr(\text{toxicity})$, so it is more desirable to have larger ξ_1 , smaller ξ_2 or both. An equivalent formulation may be constructed in terms of $\xi_2 = \Pr(\text{no toxicity})$ rather than $\Pr(\text{toxicity})$, in order to make larger values more desirable for both parameters, if desired.

Dimension reduction is central to statistical thinking. The usual motivation for computing a statistic is that it estimates an important parameter of a probability model for the data. The principle of sufficiency, that for the purpose of making inferences under a given probability model no information is lost by reducing the raw data to the sufficient statistic, provides an intuitively appealing rationale for the statistical process. A common example is that in which a data set $D_n = \{Y_1, \dots, Y_n\}$ of 1-dimensional real-valued observations is reduced to the sample mean and variance, $(\bar{Y}, s^2)$, which estimate the corresponding population values, μ and σ^2 . This relies on the idea that these two parameters are the most interesting properties of the population from which the data arose. Given this, the mapping $D_n \rightarrow (\bar{Y}, s^2)$ makes sense. In the regimes described here, the analogous dimension reduction $\boldsymbol{\theta} \rightarrow \boldsymbol{\xi}$ is done on the parameter vector itself, based on the idea that $\boldsymbol{\xi}$ characterizes what is most interesting about $\boldsymbol{\theta}$.

As a numerical example, suppose that the null historical mean outcome probabilities with standard treatment are $\xi_0 = (\xi_{0,1}, \xi_{0,2}) = (.40, .30)$ so that, considering each parameter individually, either a value of $\xi_1 > .40$ or a value of $\xi_2 < .30$ is an improvement over the historical mean. In the single-arm phase II setting, Bryant and Day (1995) proposed an optimal 2-stage frequentist test-based design that requires an alternative hypothesis with improvements in both entries of ξ_0 , such as $\{\xi: \xi_1 \geq .55, \xi_2 \leq .20\}$ in the example. This type of goal, which aims for both an increase of .15 or more in the response rate ξ_1 and a decrease of .10 or more in the toxicity rate ξ_2 compared to ξ_0 , is rather optimistic in most clinical settings, however. A more general approach with rectangular alternatives allows one of the two parameters to be equal to or worse than the null value so that, for example, the alternative might be $\{\xi: \xi_1 \geq .55, \xi_2 \leq .40\}$. This allows a .10 increase in the toxicity probability as a trade-off for a .15 increase in the probability of response. General tests for this type of alternative are described by Jennison and Turnbull (1993, 2000).

To construct a Bayesian design, a tractable model is obtained by assuming that θ follows a conjugate Dirichlet prior $p(\theta|\alpha) \propto \theta_{1,1}^{\alpha_{11}-1} \theta_{1,0}^{\alpha_{10}-1} \theta_{0,1}^{\alpha_{01}-1} \theta_{0,0}^{\alpha_{00}-1}$ where $\alpha = (\alpha_{1,1}, \alpha_{1,0}, \alpha_{0,1}, \alpha_{0,0})$ denotes a vector of positive-valued hyperparameters and we denote $\theta \sim \text{Dir}(\alpha)$. Under this model, for count data $\mathbf{X} = (X_{1,1}, X_{1,0}, X_{0,1}, X_{0,0})$ where

$X_{a,b} = \sum_{i=1}^n Y_{i,1}^a (1 - Y_{i,1})^{1-a} Y_{i,2}^b (1 - Y_{i,2})^{1-b}$ the posterior is $\theta | \mathbf{X} \sim \text{Dir}(\alpha + \mathbf{X})$ and marginally ξ_1 and ξ_2 follow beta priors and posteriors. For a single-arm phase II clinical trial with rules corresponding to the frequentist test described above, one may stop the trial and declare the treatment to be not promising for future study if either $\Pr(\xi_1 > .55 | \text{data}) < p_{1L}$ or $\Pr(\xi_2 < .40 | \text{data}) < p_{2L}$, with the decision cut-offs p_{1L} and p_{2L} in these Bayesian stopping criteria calibrated to obtain a design in which the two rules applied together give a design with good frequentist properties. If $(.40, .30)$ is the historical mean, corresponding to the null values $\xi^{(0)}$ in the frequentist tests, one may regard these two rules as allowing, on average, an increase of .10 in $\Pr(\text{toxicity})$ as a trade-off for achieving an increase of .15 in $\Pr(\text{response})$. In general, the fixed values .55 and .40 may be replaced by random parameters corresponding to historical treatment, although this is not strictly necessary. Details for constructing multiple stopping rules of this form are given by Thall, Simon and Estey (1995).

The problem becomes much more difficult if two or more equally desirable targets are specified, such as the three pairs $(.55, .40)$, $(.60, .45)$ and $(.40, .15)$. The target $(.60, .45)$ allows a .15 increase in ξ_2 as a trade-off for an increase of .20 in ξ_1 , while the target $(.40, .15)$ allows the response rate to remain at the historical value provided that on average ξ_2 drops from .30 to .15. In such settings, constructing a frequentist test is no longer straightforward, since there no longer is a simple rectangular alternative. Obtaining multiple targets from a physician is not only easy but, in my experience, a physician asked for a single target is very likely to provide several. The problem is how to construct statistical decision rules that reflect several equally desirable 2-dimensional parameter values. In particular, some sort of dimension reduction is needed, and it is not obvious how this should be done.

In practice, the most common approach to this problem is to ignore one of the two parameters, typically by using the semantic device of calling Y_2 a “secondary outcome,” and constructing a conventional decision rule or trial design based on ξ_1 alone. If, for example, the stated goal is to achieve a 50% or higher response rate, one may use simple decision criteria such as $\Pr(\xi_1 > .50 | \text{data})$ or mean $E(\xi_1 | \text{data})$ under a Bayesian model or frequentist criteria such as the sample mean response rate $\hat{\xi}_1$ or the approximately normal test statistic $(\hat{\xi}_1 - .50)/(.50 \times .50/n)^{1/2}$. Of course, this approach has many defects, including the inability to identify treatment differences in settings where, for treatments A and B , $\xi_{1,A}$

$= \xi_{1,B}$ but $\xi_{2,A} > \xi_{2,B}$. Many clinical trial protocols that take this sort of approach include a statement that toxicity will be monitored, but do not provide explicit rules for doing so. Since critical decisions pertaining to safety monitoring are made in a purely subjective manner, the operating characteristics of such designs cannot be computed. The point is that treatment differences may occur in terms of ξ_1 alone, ξ_2 alone, or both parameters.

This paper will review some geometric methods for constructing decision rules based on 2-dimensional parameters in clinical trials. In each method, treatment effect is characterized by a parameter pair $\xi = (\xi_1, \xi_2)$, typically arising from a model of higher dimension, and a geometric representation in the parameter space constructed from a set of elicited ξ values is used as a basis for defining decision rules. These methods are described in three recent papers. Thall, Sung and Estey (2002) compare two-stage treatment strategies for patients with acute leukemia based on covariate-adjusted probabilities of efficacy and death under a generalized logistic model. Thall and Cook (2004) present a general method for dose-finding in phase I/II clinical trials that uses trade-offs between the probabilities of efficacy and toxicity as a basis for selecting doses. The geometric construct used in each of these papers is a family of contours that partitions the 2-dimensional parameter space, with the contours constructed so that all parameter pairs on a given contour are equally desirable. Thall, Wooten and Shpall (2006) present a Bayesian method for comparing treatments for rapidly fatal diseases based on the probability of response and the mean time to response among responders, computed under a competing risks regression model for the times to response and death. They construct decision rules based on posterior probabilities of a convex set of desirable ξ values. While these methods deal with a variety of data structures, medical settings and statistical goals, they share several essential properties. My goals in this paper are to provide a general framework for these methods and illustrate how each may be applied in practice.

Section 2 will provide a general notational framework for what is to follow. In section 3, methods for constructing families of trade-off contours that are used as a basis for decision making are described, and the methods are illustrated by applications to actual clinical trial designs. Section 4 describes geometric methods for evaluating and comparing treatment effects based on posterior probabilities of two-dimensional parameter sets, and a brief discussion is given in section 5.

2. A General Framework

Consider a clinical trial in which the data on each patient consist of a multivariate outcome $\mathbf{Y}$ and a vector $\mathbf{Z}$ containing some combination of covariates, treatments or doses. A Bayesian model will be assumed that consists of a likelihood $f(\mathbf{Y} | \mathbf{Z}, \boldsymbol{\theta})$ and prior $p(\boldsymbol{\theta})$, with parameter vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$ of dimension $p \geq 2$. In some cases, the Bayesian structure is not strictly necessary and only the likelihood and a consistent estimator of $\boldsymbol{\theta}$ are needed. All of the methods rely on the key assumption that, from both a scientific and a medical viewpoint, it is appropriate to base treatment evaluation and decision making on a 2-dimensional parameter $(\xi_1, \xi_2) = \xi \in \Xi \subset \mathbb{R}^2$ defined by a mapping $(\boldsymbol{\theta}, \mathbf{Z}) \rightarrow \xi$. In the settings addressed here, ξ is readily interpretable, and in my experience it often is proposed by the physicians planning the trial. In contrast, the likelihood, prior and model parameter $\boldsymbol{\theta}$ are essentially statistical objects used to make statements and decisions about ξ and $\mathbf{Z}$. In the applications discussed below, Ξ is, respectively, the triangular sub-region of $[0, 1]^2$ where $0 < \xi_1 + \xi_2 \leq 1$ for probabilities corresponding to binary outcomes, the unit square $[0, 1]^2$ for unconstrained pairs of probabilities, $[0, \infty) \times [0, 100]$ for $\xi_1 = \text{mean survival time}$ and $\xi_2 = \text{mean QOL}$, and $[0, 1] \times [0, \infty)$ for $\xi_1 = \text{probability}$ and $\xi_2 = \text{mean event time}$.

The basis for the first set of methods is the idea that a family $\mathcal{F} = \{C_\delta; \delta \in R\}$ of contours may be constructed to reflect the area expert's goals in such a way that, for each real number δ , all $\xi \in C_\delta$ are equally desirable, and $\mathcal{F}$ partitions Ξ . Formally, $\bigcup_{\delta \in R} C_\delta = \Xi$ and $C_{\delta_1} \cap C_{\delta_2} = \varnothing$ for distinct δ_1 and δ_2 . Thus, δ quantifies the desirability of all $\xi \in C_\delta$. In this sense, each contour C_δ may be regarded as an “indifference” set. Depending on the method used for constructing $\mathcal{F}$, for each real δ the contour C_δ is a curve in Ξ constructed so that it satisfies a desirability constraint expressed in terms of an explicit function of ξ and δ .

In the illustrations presented below, the partition $\mathcal{F}$ is constructed in several different ways.

In the first method, one elicits several parameter pairs $\xi_1^{(e)}, \dots, \xi_K^{(e)}$ and their corresponding desirabilities $\delta_1, \dots, \delta_K$. To implement the other methods, one elicits a set of target pairs that the physician considers equally desirable, these are used to construct a target contour C , and $\mathcal{F}$ is generated by applying a homotopy to C . In some settings, the targets may be elicited in relation to an established null or historical mean value ξ_0 . Using a graphical representation of the elicited points and contours greatly facilitates the elicitation process. In constructions where a target contour C is first constructed from elicited values, for $K = 2$ or 3 one may determine C so that it passes through all of the elicited points, whereas for $K > 3$ one may treat the elicited points like regression data and use least squares or some other curve-fitting method to determine C . Although things will be described in the context of designing a clinical trial where the experts are physicians, the methods may be applied more generally, and the purpose of the exercise may be design, decision-making or data analysis.

Assume without loss of generality that larger values of ξ_1 and smaller values of ξ_2 are more desirable, corresponding to the earlier example where $\xi_1 = \text{Pr}(\text{efficacy})$ and $\xi_2 = \text{Pr}(\text{toxicity})$, so that the lower right portion of Ξ corresponds to the most desirable ξ pairs. For each method used to construct $\mathcal{F}$, the main formal requirement is that, for each real δ , considering ξ_2 as a function of ξ_1 on C_δ , ξ_2 must increase with ξ_1 on C_δ . This ensures that if either ξ_1 is increased or ξ_2 is decreased with the other parameter held constant, then the point must move to a contour having a larger desirability. Essentially, $\delta: \Xi \rightarrow R$ is a function mapping each 2-dimensional parameter of interest into a real number that can be used as a basis for comparing treatments and making decisions.

3. Constructing A Family of Contours

3.1 Eliciting Both Targets and Desirabilities

The first method for constructing a family of contours, described in Thall, Sung and Estey (2002), requires one to elicit $K = 2$ or 3 target pairs and their corresponding numerical

desirabilities, such that $f(\xi_j^{(e)}) = \delta_j$ for each $j = 1, \dots, K$, given a suitably defined function f that determines the shape of the contours. The elicited values are used to estimate the parameters of f , and thus one may determine the contour C_δ for all given δ . The practical reason for constructing a partition $\mathcal{F}$ of Ξ in this way is to map Ξ into a set of one-dimensional δ values in such a way that, if $\delta(\xi_1) > \delta(\xi_2)$, this accurately reflects the subjective belief that ξ_1 is more desirable than ξ_2 . As noted earlier, one may use either consistent frequentist estimates or Bayesian posterior means to estimate ξ . For example, if the estimates of ξ for patients given treatments A and B are $\widehat{\xi}_A$ and $\widehat{\xi}_B$, one may compare the treatments using the estimated desirabilities $\delta(\widehat{\xi}_A) = \widehat{\delta}_A$ and $\delta(\widehat{\xi}_B) = \widehat{\delta}_B$. For many applications, a model more elaborate than the multinomial-Dirichlet given earlier may be needed if one wishes to account for regression of the ξ_j 's on treatments and patient covariates known to affect outcome.

The clinical trial that motivated this method was designed to compare several two-stage chemotherapy strategies for treating patients with acute leukemia who had very poor prognosis due to the fact that their disease had re-appeared after previously being brought into remission. Because CR is a necessary but not sufficient condition for long-term survival in this patient group, chemotherapy typically is very aggressive, and patients may die due to infections or other effects that are related to both their disease and chemotherapy regimen. Patient outcome in stage $s = 1, 2$ was characterized as $\mathbf{Y}_s = (Y_{s,1}, Y_{s,2})$ where $Y_{s,1}$ indicates complete remission (CR), $Y_{s,2}$ indicates death and $Y_{s,3} = 1 - Y_{s,1} - Y_{s,2}$ indicates that neither of these events occurred, with $\mathbf{Y}_s$ scored within one month from the start of chemotherapy in that stage. For a 2-stage strategy consisting of treatments (a, b) , the patient was first given a , and a second stage of therapy with b was given if the patient neither achieved CR nor died with a in stage 1, that is, if $Y_{1,1} = Y_{1,2} = 0$ and $Y_{1,3} = 1$. This is an example of a dynamic treatment regime (Murphy, 2003; Lavori and Dawson, 2004) since the treatment given in stage 2 depends on the stage 1 outcome. For $s = 1$ let t denote the stage 1 treatment, while for $s = 2$ let t denote the 2-stage strategy. To account for stage $s = 1, 2$, outcome $j = 1, 2$ indexing CR and death, treatment or strategy t , and covariates $\mathbf{Z}$ = (previous remission duration, age), we assumed a generalized logistic regression model of the form

$$\pi_{s,j}(t, \mathbf{Z}) = \Pr(Y_{s,j}=1|t, \mathbf{Z}) = \frac{\exp(\eta_{s,j,t,\mathbf{Z}})}{1 + \exp(\eta_{s,1,t,\mathbf{Z}}) + \exp(\eta_{s,2,t,\mathbf{Z}})}, \quad (1)$$

for $s = 1, 2$ and $j = 1, 2$, with $\pi_{s,3}(t, \mathbf{Z}) = 1 - \pi_{s,1}(t, \mathbf{Z}) - \pi_{s,2}(t, \mathbf{Z})$. The linear term $\eta_{s,j,t,\mathbf{Z}}$ was formulated to account for all effects up to second order interactions, under a Bayesian model. Additional details are given in Thall, Estey and Sung (2002).

The 2-dimensional parameters of interest were the probabilities of achieving CR or dying during the two stages of therapy with a given strategy (a, b) for a given $\mathbf{Z}$, given by

$$\xi_j(a, b, \mathbf{Z}) = \pi_{1,j}(a, \mathbf{Z}) + \pi_{1,3}(a, \mathbf{Z}) \pi_{2,j}(a, b, \mathbf{Z}) \quad (2)$$

for $j = 1, 2$. In this application, the parameter set is the triangular region $\Xi_T = \{\xi : 0 \leq \xi_1, \xi_2 \leq 1 \text{ and } 0 < \xi_1 + \xi_2 \leq 1\}$ shown in Figures 1a-c. The particular desirability function used was $\delta(\xi) = \alpha_1 \xi_1 + \alpha_2 \xi_2^{\alpha_3}$ with values of $(\alpha_1, \alpha_2, \alpha_3)$ obtained by solving the $K = 3$ equations $f(.50, .15) = 1, f(.40, .40) = 0$ and $f(.30, 0) = 0$. These equations reflected the physician's subjective assignment of $\delta = 1$ to the desirable pair $\xi_1^{(e)} = (.50, .15)$ corresponding to a 50% CR rate coupled with a 15% death rate, and $\delta = 0$ to the less desirable pairs $\xi_2^{(e)} = (.40, .40)$ and $\xi_3^{(e)} = (.30, 0)$. This yielded solution $\alpha_1 = 3.333, \alpha_2 = -2.548$ and $\alpha_3 = .707$. The contour of ξ pairs with desirability δ is $C_\delta = \{\xi \in \Xi_T : f(\xi) = 3.333\xi_1 - 2.548\xi_2^{.707} = \delta\}$. Figure 1a illustrates the family $\mathcal{F} = \{C_\delta : \delta \in R\}$ used in this application. Since $\xi(a, b, \mathbf{Z})$ is a function of treatment strategy and covariates including treatment-covariate interactions, under a Bayesian model $\delta(E\{\xi(a, b, \mathbf{Z}) | \text{data}\})$ was used to determine the desirability of each 2-stage strategy for each $\mathbf{Z}$. The targets and desirabilities for the leukemia trial illustrated in Figure 1a were elicited to correspond to average historical values using this particular type of 2-stage strategy. Alternatively, the elicitation process may be carried out in terms of a reference covariate vector corresponding to a "typical" patient chosen by the physician.

Figures 1b and 1c illustrate some other possibilities. Figure 1b shows the contours for $\xi_1^{(e)} = (.50, .15)$ having $\delta_1 = 1$ and $\xi_2^{(e)} = (.40, .40)$ having $\delta_2 = 0$ in the case $K = 2$, where the solution to the linear function $\delta(\xi) = a_1\xi_1 + a_2\xi_2$ is $(a_1, a_2) = (2.8571, -2.8571)$. The linear contours in Figure 1b also correspond to the conventional approach of eliciting utilities $\mathbf{u} = (u_1, u_2, u_3)$ for the three outcomes and identifying δ with the expected utility, so that $\delta = \xi_1 u_1 + \xi_2 u_2 + \xi_3 u_3 = \xi_1(u_1 - u_3) + \xi_2(u_2 - u_3) + u_3$. This corresponds to $a_1 = u_1 - u_3$ and $a_2 = u_2 - u_3$. For example, if the utility of the outcome “neither CR nor death” is $u_3 = 0$ then CR has utility $u_1 = a_1 = 2.8571$ and death has utility $u_3 = -2.8571$, whereas if $u_3 = -1$ then $u_1 = 1.8571$ and $u_3 = -3.8571$. In Figure 1c, considering ξ_2 as a function of ξ_1 the contours are concave rather than convex, and f is characterized by $(a_1, a_2, a_3) = (3.3333, 6.6273, 1.75)$. The practical difference between the three families of contours is illustrated by considering the relationship between $(.30, 0)$ and $(.50, .15)$. In Figure 1a, which is based on the actual elicited values, these two pairs are equally desirable. In Figure 1b, $(.50, .15)$ is slightly more desirable than $(.30, 0)$. In Figure 1c, $(.50, .15)$ is much more desirable than $(.30, 0)$. In all three illustrations, ξ_2 increases as a function of ξ_1 , which ensures that increasing ξ_1 for ξ_2 fixed increases δ and increasing ξ_2 for ξ_1 fixed decreases δ .

3.2 Eliciting Equally Desirable Targets

The second and third methods for constructing contours, given in Thall and Cook (2004) and Thall, Cook and Estey (2006) use a slightly different heuristic procedure for eliciting target pairs. For these methods, Ξ is either $[0, 1]^2$ or the triangular subset of $[0, 1]^2$ where $0 \leq \xi_1 + \xi_2 \leq 1$. The first pair $\xi_1^{(e)} = (\xi_{1,1}^{(e)}, \xi_{1,2}^{(e)})$ is a desirable target specified by the physician in the interior of Ξ . After specifying $\xi_1^{(e)}$, the physician is asked for the smallest response probability that would make the second target equally desirable if toxicity were impossible, so that $\xi_2^{(e)} = (\xi_{2,1}^{(e)}, 0)$. Similarly, the physician is then asked for the largest toxicity probability that would make the third target as desirable as the first two if the response probability were maximal, so that either $\xi_3^{(e)} = (1, \xi_{3,2}^{(e)})$ if $\Xi = [0, 1]^2$ or $\xi_3^{(e)} = (1 - \xi_{3,2}^{(e)}, \xi_{3,2}^{(e)})$ in the trinary outcome case where $\xi_1 + \xi_2 \leq 1$. A target contour C is then fit to these three points, C is assigned a convenient value of δ , such as 0 or 1, and the other contours in $\mathcal{F}$ are generated by applying a homotopy that generates a contour $C_\delta \subset \Xi$ from C for given $\delta \in R$. This simplifies the elicitation process in that the physician is not asked to specify numerical desirabilities, but rather only equally desirable target probability pairs.

To implement these methods, a flexible 3-parameter function, such as

$\xi_2 = \tau(\xi_1) = \alpha_1 + \alpha_2 \xi_1^{-1} + \alpha_3 \xi_1^{-2}$, is fit to the three elicited pairs. After solving for the parameters that characterize τ , the target contour is defined as $C = \{\xi: \xi_2 = \tau(\xi_1)\} \cap \Xi$. The only general requirement is that $d\xi_2/d\xi_1 > 0$ for all ξ_1 in the subset of $[0, 1]$ where $\xi_2(\xi_1) \in \Xi$. Once C is established, to generate $\mathcal{F}$, for any pair $\mathbf{q} \in \Xi$, let $L(\mathbf{q})$ denote the straight line passing through $\mathbf{q}$ and the ideal point $(1, 0)$, so that $L(\mathbf{q})$ consists of the points in Ξ satisfying $\xi_2 = \{q_2/(1 - q_1)\}(1 - \xi_1)$. Let $\mathbf{r} = \mathbf{r}_c(\mathbf{q}) = L(\mathbf{q}) \cap C$ be the point where $L(\mathbf{q})$ and C intersect. This is illustrated in Figure 2, which corresponds to an application that will be discussed below. In Figure 2, the three elicited equally desirable targets are shown as round points, C is given by the thick line and $\mathbf{q}$ and $\mathbf{r}$ as described are given by triangular points. Denoting the Euclidean distance from any $\xi \in \Xi$ to $(1, 0)$ by $\|\xi - (1, 0)\| = \{(\xi_1 - 1)^2 + (\xi_2 - 0)^2\}^{1/2}$, the desirability of $\mathbf{q}$ is defined to be

$$\delta(\mathbf{q}) = \frac{\|\mathbf{r}_c(\mathbf{q}) - (1, 0)\|}{\|\mathbf{q} - (1, 0)\|} - 1. \quad (3)$$

Since $\mathbf{r}_c(\mathbf{q}) = \mathbf{q}$ for all $\mathbf{q} \in C$, as a convenience the number 1 is subtracted from the ratio of norms in (3) in order to make the desirability of any point on C equal to 0. Thus, points inside the contour and hence closer to (1,0) have desirability $\delta > 0$ and points outside C have $\delta < 0$. Given $\mathbf{q}$, the point $\mathbf{r}_c(\mathbf{q})$ is found by solving the two equations $\xi_2 = \tau(\xi_1)$ and $\xi_2 = \{q_2 / (1 - q_1)\}(1 - \xi_1)$. Given δ , the contour C_δ of points having this desirability is determined by the continuous function $f_\delta: C \rightarrow \Xi$ defined by

$$f_\delta(\xi) = \left(1 - \frac{1 - \xi_1}{\delta + 1}, \frac{\xi_2}{\delta + 1}\right), \quad (4)$$

which maps each $\xi \in C$ to a point such that $\|\xi - (1, 0)\| / \|f_\delta(\xi) - (1, 0)\| = \delta + 1$. homotopy Thus, $C_\delta = \{f_\delta(\xi): \xi \in C\} \cap \Xi$ for given desirability δ , and the target contour is $C = C_0$. Defining the family of contours to be $\mathcal{F} = \{C_\delta: \delta \in R\}$, it follows that $C_{\delta_1} \cap C_{\delta_2} = \emptyset$ for distinct δ_1 and δ_2 and $\bigcup_{\delta \in R} \{C_\delta: \delta \in R\} = \Xi$ so that $\mathcal{F}$ partitions Ξ .

3.3 Using An L^p Norm With Equally Desirable Targets

The third method also is based on three equally desirable elicited points, but it uses an L^p norm to define C . For $p > 0$ and $\xi \in \Xi$, define the L^p norm distance from ξ to (1,0) to be

$$\|\xi - (1, 0)\|_p = \left\{ \left(\frac{\xi_1 - 1}{\xi_{2,1}^{(e)} - 1} \right)^p + \left(\frac{\xi_2}{\xi_{3,2}^{(e)}} \right)^p \right\}^{1/p}. \quad (5)$$

In particular, the norm is defined so that $\|\xi_2^{(e)} - (1, 0)\|_p = \|\xi_3^{(e)} - (1, 0)\|_p = 1$. Setting $\delta = 0$ for points on C as before, the contour with desirability δ is defined to be

$$C_\delta = \{\xi: \|\xi - (1, 0)\|_p = 1 - \delta\} \cap \Xi.$$

It is easy to show (Cook, 2006) that, given $\xi_2^{(e)}$ and $\xi_3^{(e)}$, p is uniquely determined by $\xi_1^{(e)}$ and may be found by a monotone search. The L^p norm method has the advantage that a solution always exists, whereas when using a 3-parameter function to construct contours there is the possibility that, for particular elicited targets, the admissibility requirement of monotonicity for ξ_2 as a function of ξ_1 may be violated. Another technical difference is that, if ξ is given and one wishes to solve for its desirability δ , the method based on $\xi_2 = \alpha_1 + \alpha_2 \xi_1^{-1} + \alpha_3 \xi_1^{-2}$ requires solving a cubic equation in δ , whereas the L^p norm method gives the solution from the simple equation $\delta = 1 - \|\xi - (1, 0)\|_p$.

These methods will be illustrated with a clinical trial in which the scientific goal is to determine the best dose of tissue plasminogen activator (tPA) administered intra-arterially to children who have suffered a stroke within the past 3 hours. The therapeutic goal is to

dissolve the blood clot that caused the stroke and thus restore blood flow to the stroke site in the brain, that is, achieve “recanalization.” Response is defined as recanalization based on comparison of pre-infusion and immediate post-infusion angiograms. Toxicity is defined as death, intracranial hemorrhage or any other SAE within 48 hours of administering the tPA. The three target probability pairs considered equally desirable by the physician planning the trial were $\xi^{(1)} = (.50, .05)$, $\xi^{(2)} = (.40, 0)$, and $\xi^{(3)} = (1, .20)$. This says that the physician considered 50% response with a 5% toxicity rate, certain response with a 20% toxicity rate, and 40% response with toxicity impossible to be equally desirable.

To obtain contours using the L^p -norm method, we first use $\xi^{(2)}$ and $\xi^{(3)}$ to define

$$\|\xi - (1, 0)\|_p = \left\{ \left(\frac{\xi_1 - 1}{.40 - 1} \right)^p + \left(\frac{\xi_2 - 0}{.20 - 0} \right)^p \right\}^{1/p}.$$

The equation $\|(0.50, .05) - (1, 0)\|_p = 1$ has unique solution $p = 1.1829$, the resulting contours are given by

$$C_\delta = \{\xi \in [0, 1]^2 : \|\xi - (1, 0)\|_{1.1829} = 1 - \delta\},$$

and the desirability of a given parameter pair is $\delta(\xi) = 1 - \|\xi - (1, 0)\|_{1.1829}$. The contours are illustrated in Figure 2, which was described earlier. Alternatively, using the quadratic function $\xi_2 = \alpha_1 + \alpha_2 \xi_1^{-1} + \alpha_3 \xi_1^{-2}$ rather than the L^p norm to define C , the three elicited pairs give the solution $\alpha_1 = .4167$, $\alpha_2 = -.25$, $\alpha_3 = .0333$ so that, given δ ,

$$C_\delta = \left\{ \left(1 - \frac{1 - \xi_1}{\delta + 1}, \frac{.4167 - .25/\xi_1 + .0333/\xi_1^2}{\delta + 1} \right) : 0 \leq \xi_1 \leq 1 \right\} \cap [0, 1]^2.$$

The contours obtained using the quadratic function are virtually identical in appearance to those in Figure 2 obtained using the L^p norm method, hence are not shown.

In the tPA trial, the primary objective is to determine an optimal dose among the four levels $\{0.6, 0.8, 1.0, 1.2\}$ mg/kg tPA. Let Z denote the standardized dose obtained by taking logs and centering at the mean. The method relies on the underlying dose-outcome model given by marginals $\xi_1(Z) = \Pr(Y_1 = 1) = \Pr(\text{response}) = \text{logit}^{-1}(\beta_{1,0} + \beta_{1,1}Z + \beta_{1,2}Z^2)$ and $\xi_2(Z) = \Pr(Y_2 = 1) = \Pr(\text{toxicity}) = \text{logit}^{-1}(\beta_{2,0} + \beta_{2,1}Z)$ with joint probabilities $\Pr(\mathbf{Y} = (a, b) | Z)$ for $a, b \in \{0, 1\}$ determined by these marginals and one association parameter ψ under a Gumbel model. Thus, the model parameter vector is $\theta = (\beta_{1,0}, \beta_{1,1}, \beta_{1,2}, \beta_{2,0}, \beta_{2,1}, \psi)$, and the mapping $(\theta, Z) \rightarrow \xi(Z) = (\xi_1(Z), \xi_2(Z))$ characterizes (θ, Z) in terms of the pair of outcome probabilities as functions of dose. Additional details of the general model and method are given in Thall and Cook (2004). The design specifies a maximum of 24 patients to be treated in cohorts of size 2, starting at the lowest dose and not skipping an untried dose when escalating. Each new cohort is given the dose having the largest posterior mean $E\{\xi(Z) | \text{data}\}$ based on the most recent data from the trial, and this criterion determines the final recommended dose based on the data from all 24 patients. An important aspect of the trade-off contours (Figure 1a) that are the basis for choosing among the doses in this trial is that a small increase in $\xi_2(Z) = \Pr(\text{toxicity} | Z)$ is allowed for a relatively large increase in $\xi_1(Z) = \Pr(\text{recanalization} | Z)$ in order to maintain the same desirability.

4. Basing Decisions On Posterior Probabilities

The next method is different from those discussed above in that C is now used to define a set $\Xi_C \subset \Xi$ of parameter pairs such that each $\xi \in \Xi_C$ is at least as desirable as a pair on C , and decisions are based on posterior probabilities $\Pr(\Xi_C | \text{data})$. Thus, this method requires a Bayesian model. For an application where ξ is defined so that larger values of both parameters are more desirable, once C has been established the target set is defined to be

$$\Xi_C = \{\xi: \xi_1 \geq \xi_1^* \text{ and } \xi_2 \geq \xi_2^* \exists \xi^* \in C\} \quad (6)$$

and Ξ_C is oriented toward the upper right portion of Ξ . If the parameters are defined so that larger ξ_1 and smaller ξ_2 are more desirable than

$$\Xi_C = \{\xi: \xi_1 \geq \xi_1^* \text{ and } \xi_2 \leq \xi_2^* \exists \xi^* \in C\} \quad (7)$$

and Ξ_C is oriented toward the lower right portion of Ξ . This method allows a variety of general forms of C , including polygonal boundaries. This is illustrated by Figure 3, which corresponds to the first application that we will discuss below.

In the methods discussed in section 3, Ξ was a set of probability pairs, which facilitated the use of Euclidean or L^p norm distance to define a homotopy that was used to generate $\mathcal{F}$ from either one or two contours. In the present regime, the parameters ξ_1 and ξ_2 may be defined on different domains, as in the illustration below where ξ_1 is a probability and ξ_2 is a mean time. To construct C , one may use any of the methods described previously, define a polygonal boundary as in Conaway and Petroni (1995) or Thall and Cheng (1999), or apply the following method. Given $K \geq 2$ elicited target parameter pairs, $\xi_1^{(e)}, \dots, \xi_K^{(e)}$, the basic idea is to treat these like regression data, fit a curve to these values using some standard method such as least squares, and define C to be the fitted curve. That is, the $\xi_{j,1}^{(e)}$ s are treated as predictors and the $\xi_{j,2}^{(e)}$ s as outcome variables in a parametric regression model, such as $\xi_2 = \tau(\xi_1) = \alpha_1 + \alpha_2 \xi_1 + \alpha_3 \xi_1^2$ or some other functional form. If desired, one may use least absolute value regression, that is, solve for the parameters that minimize $|\xi_2 - \tau(\xi_1)|$, so that ξ_1 and ξ_2 are treated more symmetrically. Denoting the fitted function by $\widehat{\tau}(\xi_1)$ and the interval of ξ_1 for which $(\xi_1, \tau(\xi_1)) \in \Xi_C$ by $[\xi_{1,L}, \xi_{1,U}]$, the target contour is

$$C = \{(\xi_1, \widehat{\tau}(\xi_1)): \xi_{1,L} \leq \xi_1 \leq \xi_{1,U}\} \cap \Xi.$$

The physician may wish to modify this contour by imposing a lower bound on ξ_1 so that the left-hand portion of C is a vertical line, as in Figure 3, or similarly bound ξ_2 . While this type of construction is more flexible than the previous methods, it requires some care in implementation. Showing the physician a graph of the scattergram of elicited points and the fitted curve is essential, since this allows the physician to adjust any point that disagrees substantially with the others.

As a first illustration of this method, consider a hypothetical clinical setting based on $Y_1 =$ survival time and $Y_2 =$ quality of life (QOL), with $\xi_j = E(Y_j)$ for $j = 1, 2$. Thus, $\xi_1 =$ mean survival time and $\xi_2 =$ mean QOL, and larger values of both parameters are more desirable.

Assume that QOL is scored on a 100-point scale where 0 corresponds to the worst possible QOL, $80 \leq \text{QOL} \leq 90$ is a typical range for people who are disease-free, and 100 corresponds to perfect health and psychological well-being. In practice, as in the previous applications, these parameters typically will arise from an underlying model accounting for treatment and covariate effects, and (Y_1, Y_2) may be part of an observed vector of higher dimension. For simplicity, such additional structure will be ignored in this example. Suppose that, based on historical data, the means are $\xi_{1,0} = 3$ years and $\xi_{2,0} = 50$ with standard therapy. Given $\xi_0 = (3, 50)$, several hypothetical target sets will be described where C is constructed from three elicited pairs. The first pair $\xi_1^{(e)} = (\xi_1, \bar{\xi}_2)$ corresponds to a case where a decrease in ξ_1 from 3 years to ξ_1 is a trade-off for a very high QOL $\bar{\xi}_2$. The second pair $\xi_2^{(e)}$ corresponds to an improvement in both parameters. The third pair is $\xi_3^{(e)} = (\infty, \xi_2)$, so that ξ_2 is the lowest QOL that would be acceptable regardless of how large survival time may be. Since the parameters are defined so that larger values of both ξ_1 and ξ_2 are more desirable, Ξ_C is oriented toward the upper right of the domain $\Xi = [0, \infty) \times [0, 100]$. A simple function to determine C is the hyperbolic form $\xi_2 = (\bar{\xi}_2 - \xi_2)(\xi_1/\xi_1)^{\alpha} + \xi_2$ with $\alpha > 0$. The numerical values of the three elicited pairs are summarized in Table 1, and the target sets are illustrated in Figure 3. The target set in the upper left plot allows a decrease in mean survival from 3 to 2.5 years as a trade-off for perfect QOL, and a small drop in QOL from 50 to 45 if survival time is very large. The target in the upper right is slightly more optimistic in that the lower bound on QOL is maintained at the historical mean of 50. The set in the lower left requires a QOL of 90 as a trade-off for a 1 year drop in survival, from 3 to 2 years. The set in the lower right is the most optimistic, demanding a QOL no smaller than 70, along with a perfect QOL of 100 if survival drops from 3 to 2 years. For a given Ξ_C , a test would conclude the experimental treatment is superior to standard therapy for which $E(\xi) = (3, 50)$ if $\Pr(\Xi_C | \text{data}) > p_U$. The decision cut-off p_U and sample size could be calibrated so that, for example, the frequentist probability of concluding that the experimental treatment is superior to the standard is at most .05 when in fact $\xi = (3, 50)$ and at least .80 for all $\xi \in C$. Making decisions in this way may be contrasted with conventional tests based on rectangular alternative hypotheses, which cannot accommodate multiple equally desirable target pairs. Also, this particular application of the geometric structure provides an alternative approach to quality-adjusted survival analysis (Gelber, Gelman and Goldhirsch, 1989; Cole, Gelber and Anderson, 1994).

A clinical trial design given in Thall, Wooten and Shpall (2006) will provide a final illustration. In allogeneic transplantation (allotx) to treat hematologic malignancies, the patient's marrow is first ablated using high dose chemotherapy to destroy all the cancer cells. Bone marrow or peripheral blood stem cells from an HLA matched donor are then given to the patient, with the aim that the transplanted cells will engraft and restore the patient's bone marrow. Engraftment occurs when an absolute neutrophil count (ANC) is achieved in the patient's circulating blood at a level indicating that the transplanted cells have begun functioning normally in the bone marrow. Because a low ANC carries a high risk of fatal infection, engraftment is an essential precursor to long term survival and ideally should be achieved as soon as possible. When a matched donor cannot be found, an alternative is to use umbilical cord blood. The major disadvantage of cord blood allotx is that fewer cells are available, so it takes longer time to achieve engraftment, about 4 weeks compared to about 2 weeks with conventional allotx. A common approach to this problem is to grow more cells from those obtained initially ("ex vivo expansion") before performing the transplant.

In the trial, all patients receive a double cord blood allotx, including cells from two different placentae. Patients are randomized to receive either cells from two normal cords, or from one normal cord and one in which the cells have been expanded. Engraftment is defined as

an ANC > 500 cells per mm³ within 6 weeks. Under a competing risks model for the times to engraftment and death without engraftment, accounting for age as a covariate, the two parameters are $\xi_1 = \pi = \text{Pr}(\text{engraftment})$ and $\xi_2 = \mu = \text{mean time to engraftment among patients who engraft}$, both defined at the reference age of 38 years. Thus, larger ξ_1 and smaller ξ_2 are more desirable. The motivation for basing decisions on $\xi = (\pi, \mu)$ is that using either parameter alone ignores critical information. In comparison with the historical means $(\pi_0, \mu_0) = (.69, 30)$, the six equally desirable target pairs given by the round points in Figure 3 were elicited from the physician. The curved portion of C in Figure 3 is a simple quadratic function $\mu = \alpha_1 + \alpha_2\pi + \alpha_3\pi^2$ fit to the six points, and the requirement that $\pi \geq .60$ necessitated truncating C on the left as shown.

Denote the parameter pairs in the two treatment arms by $\xi_j = (\pi_j, \mu_j)$ for $j = 1, 2$ and let $\xi_H = (\pi_H, \mu_H)$ be corresponding historical values, so that $E(\pi_H, \mu_H) = (\pi_0, \mu_0)$. Define the pairs of differences $\Delta_{12} = (\pi_1 - \pi_2, \mu_1 - \mu_2) = -\Delta_{21}$ and $\Delta_{jH} = (\pi_j - \pi_H, \mu_j - \mu_H)$ for $j = 1, 2$. The decision rules in the trial are to terminate arm j if

$$\Pr\{\Delta_{jH} \in \Xi_c - (\pi_0, \mu_0) | \text{data}\} < p_L \quad (8)$$

and, if neither arm is terminated early, select treatment 1 at the end of the trial if

$$\Pr\{\Delta_{12} \in \Xi_c - (\pi_0, \mu_0) | \text{data}\} > \Pr\{\Delta_{21} \in \Xi_c - (\pi_0, \mu_0) | \text{data}\}, \quad (9)$$

with treatment 2 selected if this inequality holds with the indices 1 and 2 reversed. This formulation accommodates the differences Δ_{1H} , Δ_{2H} and Δ_{12} by first centering the target set, so that $\Xi_c - (\pi_0, \mu_0)$ is an improvement over $(0, 0)$ in the same way that Ξ_c is an improvement over (π_0, μ_0) . Since large $\Pr\{\xi_j \in \Xi_c | \text{data}\}$ corresponds to superiority of treatment j over the historical mean then, for example, large $\Pr\{\Delta_{12} \in \Xi_c - (\pi_0, \mu_0) | \text{data}\}$ corresponds to superiority of treatment 1 over treatment 2. Details of how the prior and p_L were calibrated to obtain a design with good properties are given in Thall, Wooten and Shpall (2006). If a specified improvement p^* on the probability domain is desired in order to conclude that one treatment is superior to the other, analogous to a test of hypothesis, then a more stringent version of rule (9) may be applied with p^* added to the right-hand side posterior probability.

5. Discussion

The methods discussed here were motivated by the desire to account for trade-offs between desirable and undesirable clinical outcomes in a way that made sense to physicians who asked for clinical trial designs dealing with such outcomes. The range of possible application seems much more general, although this type of regime has not been applied by the author outside medical statistics. The methodology as described here leads to many possible questions, including how the geometric constructs may be extended to accommodate ξ of dimension > 2 , how one may use this sort of structure to construct tests with given frequentist properties, and how such tests may compare with existing frequentist tests based on multiple outcomes.

Acknowledgments

This research was partially supported by NCI grant RO1 CA 83932. The author is grateful to Harry Whelan for giving his permission to use the IA-tPA trial as an illustration, to Xuemei Wang for assisting with preparation of the figures, and to two referees for their constructive comments.

References

- Bryant J, Day R. Incorporating toxicity considerations into the design of two-stage phase II clinical trials. *Biometrics* 1995;51:1372–1383. [PubMed: 8589229]
- Cole BF, Gelber RD, Anderson KM. Parametric approaches to quality-adjusted survival analysis. *Biometrics* 1994;50:621–631. [PubMed: 7981389]
- Conaway MR, Petroni GR. Bivariate sequential designs for phase II trials. *Biometrics* 1995;51:656–664. [PubMed: 7662852]
- Cook, JD. Technical Report UTMDABTR-003-06. Dept. of Biostatistics, M.D. Anderson Cancer Center; 2006. Efficacy-toxicity trade-offs based on L^p norms.
- Gelber RD, Gelman RS, Goldhirsch A. A quality-of-life-oriented endpoint for comparing therapies. *Biometrics* 1989;45:781–795. [PubMed: 2790121]
- Jennison C, Turnbull BW. Group sequential tests for bivariate response: Interim analyses of clinical trials with both efficacy and safety endpoints. *Biometrics* 1993;49:741–752. [PubMed: 8241370]
- Jennison, C.; Turnbull, BW. *Group Sequential Methods With Applications to Clinical Trials*. New York: Chapman and Hall; 2000.
- Lavori PW, Dawson R. Dynamic treatment regimes: practical design considerations. *Clinical Trials* 2004;1:9–20. [PubMed: 16281458]
- Murphy SA. Optimal dynamic treatment regimes (with discussion). *Journal of the Royal Statistical Society, Series B* 2003;65:331–366.
- Thall PF, Cheng SC. Treatment comparisons based on two-dimensional safety and efficacy alternatives in oncology trials. *Biometrics* 1999;55:746–753. [PubMed: 11315002]
- Thall PF, Cook John D. Dose-finding based on efficacy-toxicity trade-offs. *Biometrics* 2004;60:684–693. [PubMed: 15339291]
- Thall PF, Cook John D, Estey EH. Adaptive dose selection using efficacy-toxicity trade-offs: illustrations and practical considerations. *J Biopharmaceutical Statistics* 2006;16:623–638.
- Thall PF, Simon RM, Shen Y. Approximate Bayesian evaluation of multiple treatment effects. *Biometrics* 2000;56:213–219. [PubMed: 10783798]
- Thall PF, Sung HG, Estey EH. Selecting therapeutic strategies based on efficacy and death in multi-course clinical trials. *J American Statistical Association* 2002;97:29–39.
- Thall PF, Wooten LH, Shpall EJ. A geometric approach to comparing treatments for rapidly fatal diseases. *Biometrics* 2006;62:193–201. [PubMed: 16542246]

Figure 1. Trade-off contours for the probabilities of complete remission and death in the acute leukemia trial. Figure 1a (upper left) shows the convex contours based on the actual elicited values. Figure 1b (upper right) and Figure 1c (lower left) show hypothetical families of contours that might be obtained in practice.

Figure 2.
The family of trade-off contours used for the IA-tPA dose-finding trial.

Figure 3.

Four hypothetical target sets for ξ_1 = mean survival time and ξ_2 = mean quality of life (QOL) on a 100-point scale, based on historical means of 3 years survival and 50 on the QOL scale.

Figure 4.
The target set used as a basis for decision making in the double cord blood transplantation trial.

Table 1

Four sets of elicited target pairs used to determine a desirable set Ξ_c for ζ_1 = mean survival time and ζ_2 = mean quality of life on scale of 0 to 100, as improvements compared to the historical averages $E(\zeta_1) = 3$ years and $E(\zeta_2) = 50$.

Case	$\xi_1^{(e)}$	$\xi_2^{(e)}$	$\xi_3^{(e)}$
1	(2.5, 100)	(4, 45)	(∞ , 40)
2	(2.5, 100)	(4, 55)	(∞ , 50)
3	(2.0, 90)	(4, 55)	(∞ , 50)
4	(2.0, 100)	(4, 75)	(∞ , 70)