Journal of Statistical Research *ISSN 0256 - 422 X*
2012, Vol. 46, No. 2, pp. 187-202

BAYESIAN ADAPTIVE DOSE-FINDING BASED ON EFFICACY AND TOXICITY

PETER F. THALL
Department of Biostatistics. M.D. Anderson Cancer Center, Houston, Texas 77030, USA
Email: rex@mdanderson.org

SUMMARY

A description of the conventional paradigm for early phase clinical evaluation of a new agent is given, followed by a list of this paradigm's logical and practical flaws. This is provided initially to motivate the use of phase I-II clinical trial designs. The main body of the paper consists of a review of several practical Bayesian phase I-II designs for sequentially adaptive dose-finding based on efficacy and toxicity. These include designs taking two general approaches. The first approach uses elicited efficacy–toxicity probability pair trade-offs as decision criteria. Designs accommodating bivariate binary and trinary outcomes are discussed, as well as an elaboration that uses patient covariates to choose individualized doses. The second approach uses elicited joint utilities of ordinal (efficacy, toxicity) outcomes as a decision criterion, also including adaptive randomization to improve performance. Several illustrative applications of the methods are provided.

Keywords and phrases: Adaptive design; Clinical trial; Design; Phase I-II clinical trial; Utility

1 Introduction

This article reviews several practical Bayesian designs for dose-finding based on efficacy and toxicity in early phase clinical trials. Such designs are known as "phase I-II" to reflect the fact that they are hybrids, combining what conventionally are two separate, consecutive phases in the clinical evaluation of a new agent. Phase I-II designs are relatively new in clinical trials, developed over the past 15 years. Section 2 will provide a brief, preliminary review of the conventional phase I $\rightarrow$ phase II paradigm for evaluating a new agent. This preliminary review will include a list of the conventional paradigm's logical and practical flaws, with the aim to motivate the use of phase I-II designs.

There is a large and growing literature on early phase clinical trial designs. Interested readers may refer to the articles cited here, the reviews of Zohar and Chevret (2007) and Le Tourtneau, Lee, and Siu (2010), the books by Chevret (2006) and Cheung (2011), and the bibliographies therein. The remainder of the paper will describe two general Bayesian approaches to phase I-II designs that may be used as practical tools for conducting dose-finding trials. The first approach, presented in

Section 3, is based on elicited efficacy–toxicity probability trade-offs (Thall and Russell, 1998; Thall and Cook, 2004; Thall Cook and Estey, 2006). Versions accommodating bivariate binary or trinary outcomes will be discussed, and an elaboration that uses patient covariates to choose individualized doses (Thall, Nguyen and Estey, 2008) also will be described. The second approach, presented in Section 4, is based on elicited joint utilities of bivariate ordinal or binary (efficacy, toxicity) outcomes. The design also includes outcome-adaptive randomization among nearly optimal doses to improve performance (Thall and Nguyen, 2012). Illustrative applications of the methods will be presented.

2 The Conventional Phase I $\rightarrow$ Phase II Paradigm

New agents with possible anti-disease effects in humans first are developed through a complex process of pre-clinical laboratory experiments involving manipulation and study of molecules, then cells, and then small animals or primates. If the results of such studies are sufficiently promising, evaluation of the agent then may proceed with clinical trials, which are medical experiments with human subjects. The conventional approach for clinically evaluating a new agent begins with a phase I trial to determine an acceptably safe dose, usually called the "maximum tolerable dose" (MTD), based on toxicity. In phase I, anti-disease effect ("efficacy") usually is observed but is not used by the dose-finding algorithm.

Phase I clinical trial designs had their origins in evaluation of new cytotoxic agents, generally known as "chemotherapy", for treating cancer. This seems to have motivated the pervasive idea that higher doses of any agent are more likely to be toxic and also more likely to kill cancer cells and thus be efficacious. Dose-finding methods also are used in many areas outside oncology. Some examples are fibrinolytic agents for treating stroke (Whelan, et al., 2008; Thall et al. 2011), anesthetics used in surgery (Dougherty, et al., 2000), antiemetics, and anti-hypertensive agents for high blood pressure. For the purpose of dose-finding, "dose limiting toxicity" (DLT, or simply "toxicity") usually is defined as a binary indicator that any of several specified adverse events occur. Most phase I designs are based on empirical evaluation of how toxicity rates vary with dose, and they ignore treatment efficacy. In a typical phase I trial, the starting dose is chosen by the investigator based based *in vitro* or animal data. During the trial, doses are chosen adaptively for successive cohorts, typically of 1, 2, or 3 patients. This sequential approach is taken for ethical reasons, primarily due to the fear of overdosing patients. Otherwise, such as in animal experiments, a much more informative approach would be simply to randomize subjects among doses, or possibly use adaptive rules to optimize estimation of a dose-toxicity curve (cf. Atkinson, Donev, and Tobias, 2007). Decisions to escalate or de-escalate dose levels are made using using any of a wide variety of adaptive rules. Phase I trials usually are very small, enrolling about 12 to 30 patients. This is motivated by the common belief that conducting small phase I trials is a good idea, and that it is acceptable when summarizing phase I results to ignore the basic statistical practice of quantifying estimation reliability.

After a putatively safe MTD has been determined in phase I based on toxicity, a phase II trial is conducted to decide whether the treatment's efficacy is sufficiently promising to warrant further evaluation of the agent. Most often, efficacy is characterized in phase II by a binary indicator of an

event, called "response", thought to be associated with long term patient benefit. Examples of response include $\geq$50% shrinkage of a solid tumor, complete remission of acute leukemia, successful engraftment of a stem cell transplant, lowering blood pressure by a specified amount, or dissolving the arterial blood clot that caused an acute ischemic stroke. Most phase II designs in oncology are single-arm and based only on treatment efficacy, although some designs also include toxicity (Bryant and Day, 1995; Conaway and Petroni, 1995; Thall, Simon and Estey, 1995). Randomization in phase II oncology trials remains controversial (cf. Simon, Wittes, and Ellenberg, 1985; Taylor, Braun and Li, 2006; Rubinstein, et al., 2009), although randomized phase II trials are more common outside oncology. Phase II designs usually include futility rules to stop the trial early if the interim data show that the agent is not promising (Fleming, 1982; Simon, 1989; Thall and Simon, 1994). Phase II sample sizes vary widely. If it is decided in phase II that a treatment is promising, then it is evaluated in a large, confirmatory, randomized phase III trial based on a long-term outcome, most often survival time or disease-free survival time.

This paradigm has the following logical, scientific, and ethical flaws:

(1) While efficacy almost always is recorded in phase I, ignoring it when doing dose-finding wastes valuable information on how anti-disease effects may vary with dose.

(2) Ignoring efficacy in phase I dose-finding is at odds with the fact that the primary purpose of treatment is to achieve anti-disease effect.

(3) For non-cytotoxic agents, such as targeted, cytostatic, or biologic agents, if the probability of efficacy is not monotonically increasing with dose, then the premise of higher efficacy with higher dose that implicitly underlies nearly all phase I toxicity-based methods is just plain wrong.

(4) Despite the fact that commonly used "3+3" algorithms are well known to be greatly inferior to model-based procedures such as the continual reassessment method (O'Quigley, et al., 1990; Cheung, 2011), variants of the 3+3 algorithm are used most often in phase I.

(5) Because most phase I trials are small, regardless of method, they fail to estimate the dose-toxicity probability curve reliably. The consequence is that the chosen MTD is likely to be either unacceptably toxic or ineffective. For example, if one patient in six treated at an MTD has toxicity, then from a Bayesian viewpoint assuming a beta(.25, .75) prior, the posterior 95% credible interval for Pr(toxicity at the MTD) is .01 – .51. A frequentist confidence interval is similarly very wide for such small samples. In terms of statistical reliability, to declare that "an MTD has been determined" based on such data is nonsense.

(6) Since phase II trial protocols often include provisions for informal dose adjustments to deal with unacceptably high toxicity, a phase II design based on efficacy that ignores this common practice is a fictional account of what actually is done during trial conduct.

(7) The conventional practice of evaluating efficacy and toxicity separately is at odds with the routine considerations of efficacy-toxicity trade-offs that are the basis for most physicians' decision-making in treatment of severe or life-threatening diseases.

(8) If in fact an agent at a chosen MTD is ineffective but a higher dose has a substantive anti-disease effect and also is acceptably safe, then failure to explore such an efficacious higher dose is a scientific, medical, and ethical disaster. This is closely related to point (7), above, and is one of the strongest motivations for conducting phase I-II trials.

The phase I-II designs reviewed here address all of the above flaws inherent in the conventional paradigm. These phase I-II designs share the following features:

(1) They use both efficacy and toxicity to choose doses adaptively.

(2) They account explicitly for efficacy-toxicity trade-offs, either through probability trade-off functions or utilities of joint (efficacy, toxicity) outcomes.

(3) The underlying probability models allow flexible, possibly non-monotone dose-efficacy and dose-toxicity relationships.

(4) Computer simulation is used to establish each design's properties, including sample size distributions and selection probabilities for each dose, and to calibrate design parameters on that basis.

(5) The methods are Bayesian.

3 Trade-Off-Based Designs

3.1 Bivariate Binary Outcomes

Denote efficacy by E, and toxicity by T. Let Y_E = 1 if the patient experiences efficacy, and 0 otherwise. Let Y_T be the corresponding outcome for toxicity. Denote $\boldsymbol{Y} = (Y_E, Y_T)$. Given a set of doses, $d_1 < d_2 < \cdots < d_k$, let $\mathcal{D}_n = \{(\boldsymbol{Y}_1, d_{[1]}), \cdots, (\boldsymbol{Y}_n, d_{[n]})\}$ denote the data from the first n patients in the trial. The phase I-II problem is to choose a dose for each new patient or cohort during the trial, adaptively based on the most recent data $\mathcal{D}_n$, and recommend an optimal dose at the end of the trial.

For a patient given dose d, denote $\pi_j(d, \boldsymbol{\theta}) = \Pr(Y_j = 1 \mid d, \boldsymbol{\theta})$, and denote the joint probabilities $\pi(a, b \mid d, \boldsymbol{\theta}) = \Pr(Y_E = a, Y_T = b \mid d, \boldsymbol{\theta})$, for (a, b) = (1,1), (1,0), (0,1), or (0,0), where $\boldsymbol{\theta}$ is the model parameter vector. For brevity, d or $\boldsymbol{\theta}$ will be suppressed when no meaning is lost. To specify a model, let g be a suitable link function (e.g. logit, probit, or complementary log-log), with linear terms $\eta_j = g(\pi_j)$. A model may be constructed by specifying the marginals $\pi_E = g^{-1}(\eta_E)$ and $\pi_T = g^{-1}(\eta_T)$, and assuming a copula to obtain a joint distribution for $\pi(a, b)$. There are many bivariate copulas (Nelsen, 1999). The Gumbel-Morgenstern copula is quite tractable, and is given by

$$\pi(a,b) \;=\; \pi_E^a\,(1-\pi_E)^{1-a}\,\pi_T^b\,(1-\pi_T)^{1-b} + (-1)^{a+b}\left(\frac{e^\psi - 1}{e^\psi + 1}\right)\pi_E\;(1-\pi_E)\,\pi_T\,(1-\pi_T), \quad (3.1)$$

for $a, b \in \{0, 1\}$ and real-valued association parameter ψ. To stabilize computation, a standardized dose x usually is used, e.g. $x = \log(d)$ or $x = d/d_k$, possibly centered at the mean to avoid

covariance with the constant term. A simple quadratic linear term $\eta_E = \beta_{E,0} + \beta_{E,1}x + \beta_{E,2}x^2$ is reasonably flexible and allows non-monotone dose-response. If it is appropriate to assume that $\pi_E(d, \boldsymbol{\theta})$ is increasing in d, this may be achieved by setting $\beta_{E,2} = 0$ and requiring $\beta_{E,1} > 0$. For toxicity, $\pi_T(d, \boldsymbol{\theta})$ is increasing in d, and $\eta_T = \beta_{T,0} + \beta_{T,1}x$ with $\beta_{T,1} > 0$. This gives a model with p = either 5 or 6 parameters, $\boldsymbol{\theta} = (\beta_{E,0}, \beta_{E,1}, \beta_{E,2}, \beta_{T,0}, \beta_{T,1}, \psi)$. Many other functional forms for $\pi_T(d)$ and $\pi_E(d)$ may be used (cf. Bretz, Pinheiro, and Branson, 2005).

The efficacy-toxicity trade-off method (Thall and Russell, 1998; Thall and Cook, 2004; Thall, Cook and Estey, 2006) is based on a family of contours partitioning the two-dimensional set of possible marginal outcome probability pairs, $\boldsymbol{\pi} = (\pi_E, \pi_T)$. For bivariate binary outcomes, this set is $[0, 1]^2$. For a trinary outcome where E and T are disjoint, the set of possible $\boldsymbol{\pi}$ pairs is the triangular subset $\{\boldsymbol{\pi} \in [0, 1]^2 : \pi_E + \pi_T \leq 1\}$. On each contour, the trade-off between π_E and π_T is characterized as a single number, $\delta(\boldsymbol{\pi})$, the desirability of $\boldsymbol{\pi}$. The contours are constructed by first eliciting several equally desirable target probability pairs from the physician. An target efficacy-toxicity trade-off contour, $\mathcal{C}$, is obtained by fitting a curve to the elicited target pairs such that, moving along $\mathcal{C}$, as π_T increases π_E also must increase. A family of trade-off contours partitioning $[0,\ 1]^2$ is generated from $\mathcal{C}$, and a desirability, δ, is assigned to each contour in such a way that contours closer to the ideal point (π_R, π_T) = (1,0) have larger δ and contours farther away from (1,0) have smaller δ.

There are many way to construct a family of trade-off contours from $\mathcal{C}$. A simple method is as follows. For given $\boldsymbol{\pi}$, let $\boldsymbol{\pi}_{\mathcal{C}}$ denote the point where the straight line from $\boldsymbol{\pi}$ to (1,0) intersects $\mathcal{C}$. The desirability of $\boldsymbol{\pi}$ is defined as the Euclidean distance from $\boldsymbol{\pi}$ to (1,0) divided by the Euclidean distance from $\boldsymbol{\pi}_{\mathcal{C}}$ to (1,0), formally $\delta(\boldsymbol{\pi}) = \|\boldsymbol{\pi}_{\mathcal{C}} - (1, 0)\|/\|\boldsymbol{\pi} - (1, 0)\|$. To avoid infinite values, $\delta(\boldsymbol{\pi})$ may be replaced by $1 - e^{-\delta(\boldsymbol{\pi})}$. Let $\mathcal{C}_{\delta^*} = \{\boldsymbol{\pi} :\ \delta(\boldsymbol{\pi}) = \delta^*\}$ denote the contour on which all $\boldsymbol{\pi}$ have desirability δ^*. This gives an ordering of all $\boldsymbol{\pi}$ pairs in terms of $\delta(\boldsymbol{\pi})$. To use this construction for dose-finding, note that the above construction induces desirabilities δ^* on $\{d_1, \cdots, d_k\}$ in terms of the desirabilities of their pairs of posterior means,

$$\delta^*(d_r) \ = \ \delta[E\{\pi_E(d_r, \boldsymbol{\theta}) \mid \mathcal{D}\}, E\{\pi_T(d_r, \boldsymbol{\theta}) \mid \mathcal{D}_n\}], \qquad r = 1, \cdots, k.$$

An alternative approach would be to use contours that are linear in the $\boldsymbol{\pi} = (\pi_E, \pi_T)$ domain, so that if $\delta(\boldsymbol{\pi}^{(1)}) = \delta(\boldsymbol{\pi}^{(2)})$ then all linear combinations $\lambda\boldsymbol{\pi}^{(1)} + (1 - \lambda)\boldsymbol{\pi}^{(2)}$, for $0 < \lambda < 1$, on the straight line connecting $\boldsymbol{\pi}^{(1)}$ and $\boldsymbol{\pi}^{(2)}$ also must have desirability δ. However, requiring linearity in the $\boldsymbol{\pi}$ domain does not seem to make sense since, for example, the desirabilities in the real domain $(\text{logit}(\pi_E), \text{logit}(\pi_T))$ would not be linear. An important point is that is that, in this approach, the desirabilities are assigned to pairs of parameters, not to the pairs (Y_E, Y_T) of observables. Designs that take this latter approach are discussed below, in Section 4.

Since it is possible that no dose is both acceptably safe and efficacious, the following two acceptability criteria are imposed. Given upper limit $\bar{\pi}_T$ on $\pi_T(d_r, \boldsymbol{\theta})$ and lower limit $\underline{\pi}_E$ on $\pi_E(d_r, \boldsymbol{\theta})$, a dose d_r is (1) *unacceptably inefficacious* if $\Pr\{\pi_E(d_r, \boldsymbol{\theta}) < \underline{\pi}_E | \mathcal{D}_n\} > p_{U,E}$, and (2) is *unacceptably toxic* if $\Pr\{\pi_T(d_r, \boldsymbol{\theta}) > \bar{\pi}_T | \mathcal{D}_n\} > p_{U,T}$, where $p_{U,E}$ and $p_{U,T}$ are probability cut-offs such as .90 or .95. During the trial, based on the posterior from the most recent data, only acceptable doses are administered.

Table 1: Operating characteristics of the stem cell transplantation trial design.

	Dose of L				
	25	50	75	100	None
Scenario 1					
True (π_T, π_E)	(.05, .15)	(.065, .20)	(.085, .28)	(.10, .40)	
$\pi_E - \pi_T$ Trade-off	.96	1.00	1.07	1.20	
% Selected	13	11	19	56	1
# Treated	11.5	8.3	10.9	28.9	
Scenario 2					
True (π_T, π_E)	(.05, .15)	(.065, .45)	(.085, .20)	(.10, .10)	
$\pi_E - \pi_T$ Trade-off	.96	1.33	.98	.85	
% Selected	31	40	17	8	4
# Treated	18.2	17.7	11.5	10.7	
Scenario 3					
True (π_T, π_E)	(.15, .15)	(.20, .20)	(.35, .30)	(.40, .35)	
$\pi_E - \pi_T$ Trade-off	.85	.83	.73	.70	
% Selected	61	25	0	2	12
# Treated	29.6	17.6	4.7	3.0	

To implement the methodology, a non-informative prior $p(\boldsymbol{\theta} \mid \tilde{\boldsymbol{\theta}})$ must be established. This can be done by (i) assuming each parameter (or its log if it is positive-valued) is normally distributed, $\theta_l \sim N(\tilde{\mu}_l, \tilde{\sigma}_l^2)$, (ii) eliciting $2k$ prior means $\mu_{j,r}^{(e)}$ of $\pi_j(d_r, \boldsymbol{\theta})$ for $j = E, T$ and doses $r = 1, \cdots, k$, and (iii) solving for the prior mean vector $\tilde{\boldsymbol{\mu}}$. This can be done by treating the $2k$ elicited means like outcomes and E$(\pi_j(d_r, \boldsymbol{\theta}) \mid \tilde{\boldsymbol{\theta}})$ as a function of $\tilde{\boldsymbol{\mu}}^{k \times 1}$ and using nonlinear least squares (Thall and Cook, 2004, section 5.1). Alternatively, a sampling-based method that simulates pseudo data from the elicited means $\mu_{j,r}^{(e)}$ may be used (Houede, et al., 2010; Thall and Nguyen, 2012). With either approach, the hyper-variances $\tilde{\sigma}_r^2$ then must be calibrated to obtain a suitably small prior effective sample size (ESS) and a design with good properties. The ESS may be computed by approximating the prior of each $\pi_j(d_r, \boldsymbol{\theta})$ with a beta(a, b), and using the prior mean and variance of $\pi_j(d, \boldsymbol{\theta})$ to solve for $a + b$, as $ESS_{j,r}(\tilde{\boldsymbol{\theta}}) \approx \mu_{j,r}(\tilde{\boldsymbol{\theta}})\{1 - \mu_{j,r}(\tilde{\boldsymbol{\theta}})\}/\text{var}\{\pi_j(d_r, \boldsymbol{\theta}) \mid \tilde{\boldsymbol{\theta}}\} - 1$. The mean over j and r then gives an overall ESS. If desired, the formal method of Morita, Thall and Mueller (2010) may be used.

3.2 Illustration: A Stem Cell Transplantation Trial

A phase I/II trial was conducted using Melphalan and an experimental agent, L, as a preparative regimen for myeloma patients undergoing an autologous stem cell transplant (SCT). Each patient received a fixed dose of Melphalan (100 mg/m^2 IV on days -3 and -2) plus one of the four doses $\{25, 50, 75, 100\}$ mg of L given orally on each of days -8 to -2. The goal was to determine an optimal dose of L. Toxicity was defined as regimen-related death, graft failure, or grade 3,4 atrial fibrillation, deep venous thrombosis, or pulmonary embolism from the start of treatment to day 30 post transplant. Efficacy was the event that the patient was alive with complete response (CR) at day 30. The design specified a maximum of 60 patients to be treated in cohorts of 3, starting at $d = 25$. A dose d was unacceptably inefficacious if $\Pr\{\pi_E(d, \boldsymbol{\theta}) < .15 | \mathcal{D}_n\} > .90$, and had unacceptably high toxicity if $\Pr\{\pi_T(d, \boldsymbol{\theta}) > .20 | \mathcal{D}_n\} > .90$. Each cohort after the first was treated with the d having largest posterior mean efficacy-toxicity trade-off, subject to the safety constraint that no untried dose level may be skipped when escalating. The trial would be terminated and no dose selected if it was found that no dose was acceptable. The elicited prior mean marginal efficacy probabilities at the four L doses were .15, .20,.25, .30, and the elicited prior mean marginal toxicity probabilities were .05, .065, .085, .10. The trade-off contours derived from the equally desirable elicited target pairs (π_E, π_T) = (.15, 0), (.30, .15), (1, .50). The efficacy-toxicity target contour is illustrated in Figure 1. Operating characteristics of the design are summarized in Table 1.

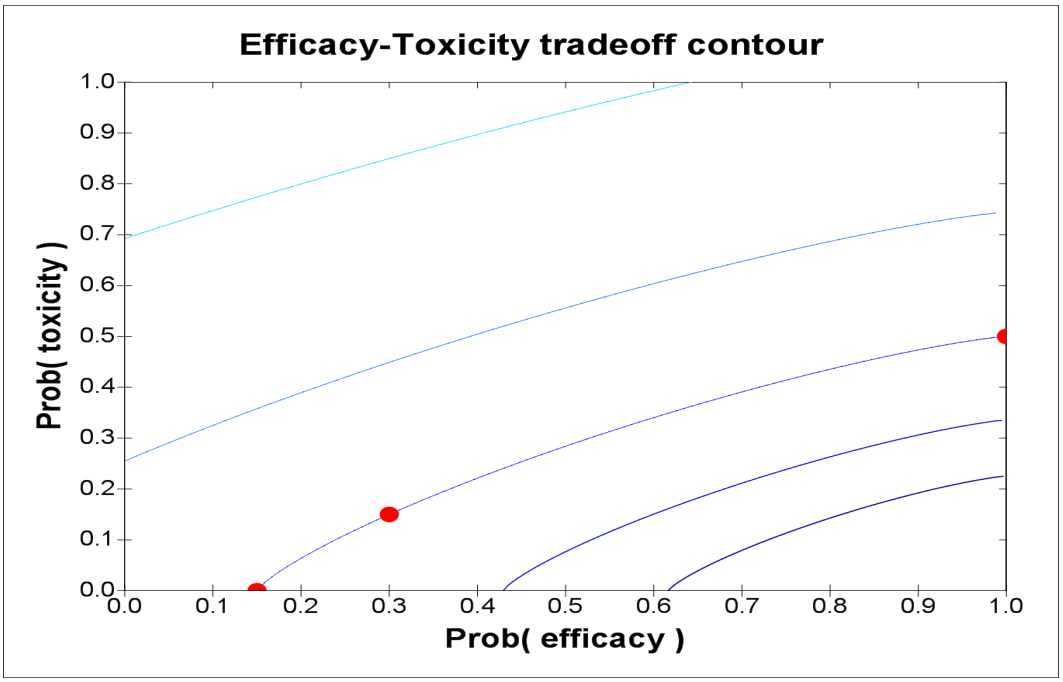

Figure 1: Trade-off contours used by the stem cell transplantation trial design.

3.3 Trinary Outcomes

The method also accommodates settings with trinary outcomes where E and T are disjoint. Denoting $N = (E \cup T)^c$, the event that neither E nor T occurs, (Y_E, Y_T) takes on the three possible values (1,0), (0,0), (0,1) corresponding to $\{E, N, T\}$, and $\{\boldsymbol{\pi} : \pi_E + \pi_T \leq 1\}$ is the triangular subset of $[0,\ 1]^2$ comprising the possible $\boldsymbol{\pi}$ pairs. This was the original case treated by Thall and Russell (1998). In this case, π_E and π_T determine all joint probabilities, with $\pi_N = 1 - \pi_E - \pi_T$. One may assume the $p = 4$ parameter continuation ratio model, with $\text{logit}^{-1}\{\pi_T(x, \boldsymbol{\theta})\} = \eta_T(x, \boldsymbol{\theta}) = \beta_{T,0} + \beta_{T,1}x$ and $\text{logit}^{-1}\{\Pr(E \mid T^c, x, \boldsymbol{\theta})\} = \eta_E(x, \boldsymbol{\theta}) = \beta_{E,0} + \beta_{E,1}x$, subject to the constraints $\beta_{T,1} > 0$ and $\beta_{E,1} > 0$. Thus $\boldsymbol{\theta} = (\beta_{E,0}, \beta_{E,1}, \beta_{T,0}, \beta_{T,1})$. This gives the three outcome probabilities

$$\pi_T(x, \boldsymbol{\theta}) = \frac{e^{\eta_T(x,\boldsymbol{\theta})}}{\{1 + e^{\eta_T(x,\boldsymbol{\theta})}\}}$$

$$\pi_E(x, \boldsymbol{\theta}) = \frac{e^{\eta_E(x,\boldsymbol{\theta})}}{\{1 + e^{\eta_T(x,\boldsymbol{\theta})}\}\{1 + e^{\eta_E(x,\boldsymbol{\theta})}\}}$$

$$\pi_N(x, \boldsymbol{\theta}) = \frac{1}{\{1 + e^{\eta_T(x,\boldsymbol{\theta})}\}\{1 + e^{\eta_E(x,\boldsymbol{\theta})}\}}$$

For this model, as x increases $\pi_T(x, \boldsymbol{\theta})$ must increase since $\beta_{T,1} > 0$, but $\pi_E(x, \boldsymbol{\theta})$ is not monotone in dose. As noted by Thall and Cook (2004) and Mandrekar, Cui and Sargent (2007), the $p = 3$ parameter proportional odds model with the same $\pi_T(d, \boldsymbol{\theta})$ but $\text{logit}^{-1}\{\pi_E(d, \boldsymbol{\theta}) + \pi_T(d, \boldsymbol{\theta})\} = \beta_{T,0} + \beta_E + \beta_{T,1}x$ with $\beta_E > 0$ used by Thall and Russell (1998) may be inadequate in this case. A four-parameter model usually is preferable, although this is not true in all cases.

As a simple illustration of the method in this case, suppose that, in the SCT trial described above, efficacy is re-defined as the patient being alive with CR at day 30, without toxicity. Thus, $T \cap E$ is empty by definition, i.e., what was previously "E and T" is now scored as T. To implement the design using this trinary outcome, one must elicit the prior, numerical acceptability criteria for $\pi_E(d, \boldsymbol{\theta})$ and $\pi_T(d, \boldsymbol{\theta})$, and trade-off contour appropriate for these outcomes. For example, if the equally desirable (π_E, π_T) = targets were (.15, 0), (.25, .10), (.65, .35), this would give the trade-off contours illustrated in Figure 2.

3.4 Individualized Dose-Finding

A very useful elaboration of this design (Thall, Nguyen and Estey, 2008; Thall and Nguyen, 2010) incorporates patient prognostic covariates, $\mathbf{Z} = (Z_1, \cdots, Z_q)$, known to have substantive effects on π_E and π_T. The linear terms are generalized to include covariate effects $\boldsymbol{\beta}_j\mathbf{Z} = \beta_{j,1}Z_1 + \cdots + \beta_{j,q}Z_q$ and dose-covariate interactions $x\boldsymbol{\gamma}_j\mathbf{Z} = x(\gamma_{j,1}Z_1 + \cdots + \gamma_{j,q}Z_q)$. For standardized dose x in the trial,

$$\eta_j(x, \mathbf{Z}, \boldsymbol{\theta}) \ = \ f_j(x, \boldsymbol{\alpha}_j) \ + \ \boldsymbol{\beta}_j\mathbf{Z} \ + \ x\boldsymbol{\gamma}_j\mathbf{Z}. \tag{3.2}$$

The function $f_j(x, \boldsymbol{\alpha}_j)$ characterizes the main dose effects on π_j, for $j = E, T$, and is chosen to reflect the particular application. E.g. for a cytotoxic agent, $f_T(x, \boldsymbol{\alpha}_T) = \alpha_{T,0} + \alpha_{T,1}x$ subject to $\alpha_{T,1} > 0$. The key point is that $\eta_j(x, \mathbf{Z}, \boldsymbol{\theta})$ includes a dose effect term $f_j(x, \boldsymbol{\alpha}_j)$, covariate main

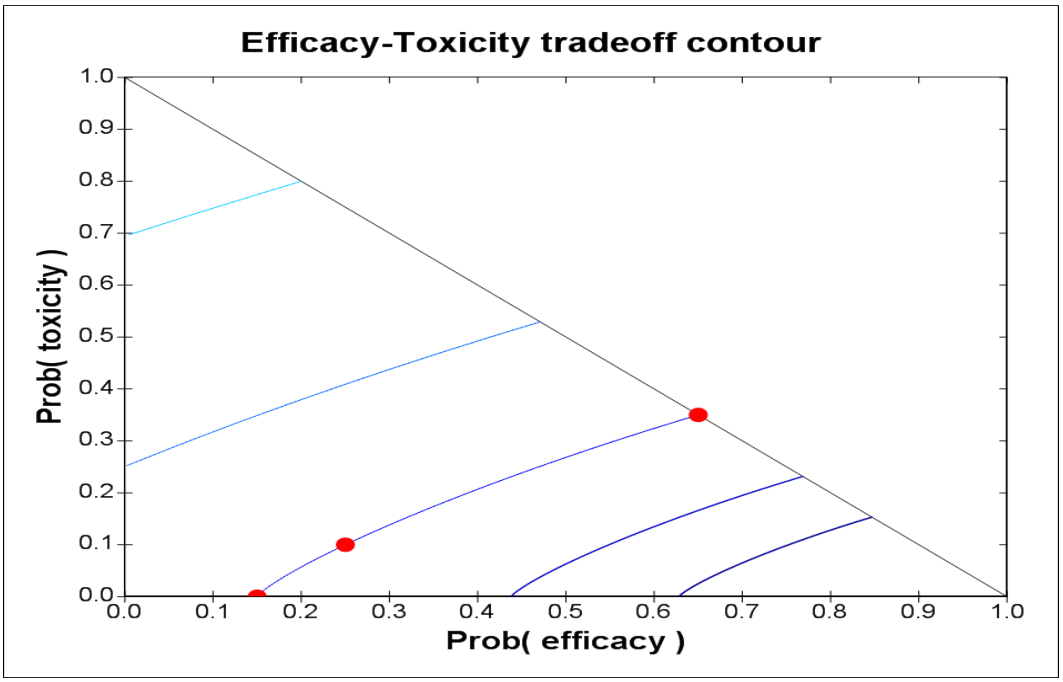

Figure 2: Trade-off contours used by the stem cell transplantation trial design with a trinary outcome.

effects, $\boldsymbol{\beta}_j\mathbf{Z}$ and dose-covariate interactions $x\boldsymbol{\gamma}_j\mathbf{Z}$. As show by Thall, Nguyen and Estey (2008), if the interaction terms are omitted, then the method may perform very poorly if such effects actually are present.

The key steps for constructing a design are as follows. First, a historical data set $\mathcal{H}$ must be fit to obtain an informative distribution on $\boldsymbol{\beta} = (\boldsymbol{\beta}_E, \boldsymbol{\beta}_T)$ and ψ, which is used as the prior on these parameters for the trial. For historical treatment $l = 1, ..., m$, it is assumed that $\eta_j(\tau_l, \mathbf{Z}, \boldsymbol{\theta}) = \mu_{j,l} + \boldsymbol{\beta}_j\mathbf{Z}$, so the covariate-adjusted historical treatment effects on Y_j are $\boldsymbol{\mu}_j = (\mu_{j,1}, \cdots, \mu_{j,m})$. The trial data are used to learn about the dose main effects $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_E, \boldsymbol{\alpha}_T)$ and dose-covariate interactions $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_E, \boldsymbol{\gamma}_T)$. For each interim decision in the trial, quantities computed from the posteriors $p(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \psi \mid \mathcal{D}_n \cup \mathcal{H})$ are used for adaptive decision making. Two criteria for choosing doses are used. The first criterion is the desirability of each $\boldsymbol{\xi} = (\pi_E, \pi_T)$, using trade-offs as before. The second criterion is used to determine whether x is *acceptable for given* $\mathbf{Z}$, and is defined as follows. A lower bounding function $\underline{\pi}_E(\mathbf{Z})$ on $\pi_E(x, \mathbf{Z}, \boldsymbol{\theta})$ and an upper bounding function $\overline{\pi}_T(\mathbf{Z})$ on $\pi_T(x, \mathbf{Z}, \boldsymbol{\theta})$ as $\mathbf{Z}$ is varied are obtained constructively from elicited values. Given $\mathcal{D}_n$, the set of acceptable doses for a patient with covariates $\mathbf{Z}$ is all x such that

$$\Pr\{\pi_E(x, \mathbf{Z}, \boldsymbol{\theta}) < \underline{\pi}_E(\mathbf{Z}) \mid \mathcal{D}_n \cup \mathcal{H}\} > p_{U,E} \text{ and } \Pr\{\pi_T(x, \mathbf{Z}, \boldsymbol{\theta}) > \overline{\pi}_T(\mathbf{Z}) \mid \mathcal{D}_n \cup \mathcal{H}\} > p_{U,T}. \tag{3.3}$$

During the trial, a patient with covariates $\mathbf{Z}$ is given the most desirable acceptable dose, or is not

Table 2: Hypothetical recommended covariate-specific doses for the AML trial. IND denotes the number of prior induction regimens.

	CR duration < 52 weeks		CR duration > 52 weeks	
Age	IND=2	IND=1	IND=2	IND=1
18 – 33	1600	2200	3000	4100
34 – 42	1600	2200	2200	3000
43 – 58	1600	2200	2200	2200
59 – 66	1100	1600	1600	2200
> 66	No Dose	1100	1600	1600

treated on protocol if no dose is desirable. The key operational difference is that patients with different covariates may receive different doses, and the final result of the trial is not one selected dose, but rather and algorithm for assigning the best x based on $\mathbf{Z}$. That is, the dose assignment is *individualized*.

This method was applied to design and conduct a phase I/II trial of a new agent for patients with advanced AML, studying seven doses $\{$ 1100, 1600, 2200, 3000, 4100, 5600, 7600$\}$ mg/m^2. The covariates were age, and binary indicators of whether the patient's number IND of previous induction regimens was one or two, and whether the previous CR duration was > 52 weeks. While space limitations do not permit a detailed account of this trial, it is worthwhile to illustrate possible trial results. Hypothetical recommended covariate-specific doses are given in Table 2, which illustrates that the final recommendations of this design are very different from choosing one optimal dose for all patients.

4 Utility Based Designs

A different approach to phase I-II dose-finding is based on elicited utilities of the possible outcome pairs (Houede, et al, 2008; Thall and Nguyen, 2012). This methodology accommodates bivariate ordinal (efficacy, toxicity) outcomes, including bivariate binary outcomes as an important special case. A simple illustrative example of joint utilities for such outcomes is given in Table 3. The method described below uses such numerical values to compute the posterior mean utility of each dose as a basis for adaptive dose selection. In Table 3 note that, for example, $U(SD,\, Mild)$ = $U(CR,\, High)$ = 75, quantifying the intuitive notion that a higher level of toxicity is an acceptable trade-off for a higher level of efficacy. Another important point is that, if a dose-finding method based on only a binary toxicity defined as $\{$High or Severe$\}$ were used, then the utility of "toxicity" could be anything from 0 to 75, while the utility of "no toxicity" could be anything from 20 to 100.

Table 3: Joint utilities for a hypothetical clinical trial to treat solid tumors with ordinal efficacy and toxicity outcomes. The efficacy levels are PD = progressive disease, SD = stable disease, PR = partial response, CR = complete response.

	Toxicity Severity			
Efficacy	Mild	Moderate	High	Severe
PD	30	20	10	0
SD	75	50	30	15
PR	90	80	50	25
CR	100	90	75	45

4.1 Probability Models

Index doses by $x \in \{1, \cdots, k\}$. Let $Y_j = 0, 1, \cdots, m_j$ identify the ordinal levels of outcome $j = E, T$. Let $Y_T = 0$ be the least severe level and m_T the most severe level of toxicity, and let $Y_E = 0$ the worst level and m_E the best level of efficacy. Denote

$$\lambda_{j,y,x} \;=\; \Pr(Y_j \geq y \mid Y_j \geq y-1, x, \boldsymbol{\theta}_j) \quad \text{and} \quad \pi_{j,y,x} \;=\; \Pr(Y_j = y \mid x, \boldsymbol{\theta}_j) \tag{4.1}$$

for $j = E, T$, $y = 1, \cdots, m_k$, and dose x. Given monotone increasing link function g, such as the logit, probit, or complementary log-log, a marginal model is determined by $g(\lambda_{j,y,x}) = \theta_{j,y,x}$, for real-valued $\theta_{j,y,x}$. Denote $\boldsymbol{\theta}_j = (\boldsymbol{\theta}_{j,1}, \cdots, \boldsymbol{\theta}_{j,m_j})$, and $\boldsymbol{\theta}_{j,y} = (\theta_{j,y,1}, \cdots, \theta_{j,y,k})$. This marginal model is saturated, with m_j parameters for each x, and $\dim(\boldsymbol{\theta}_j) = km_j$, the number of $\pi_{j,y,x}$'s needed to specify the k marginals of Y_j for all x. Denoting $\lambda_{j,m_j+1,x} \equiv 1$ for convenience, the marginal probabilities are given by

$$\pi_{j,y,x} = (1 - \lambda_{j,y+1,x}) \prod_{r=1}^{y} \lambda_{j,r,x}, \qquad y = 1, \cdots, m_j. \tag{4.2}$$

To reduce computation, obtain monotonicity of $\bar{\pi}_{j,y,x} = \Pr(Y_j \geq y \mid x, \boldsymbol{\theta}_j)$ in x, and borrow strength between doses, the model may be re-parameterized as

$$\theta_{j,y,x} = \mu_{j,y} + \sum_{z=2}^{x} \gamma_{j,y,z} \qquad \text{for all } x = 2, \cdots, k \tag{4.3}$$

for real-valued $\mu_{j,y} \equiv \theta_{j,y,1}$ and $\gamma_{j,y,x} \geq 0$ for all j, y, and $x = 2, \cdots, k$. The marginal parameter vectors are now $\boldsymbol{\theta}_{j,y} = (\mu_{j,y}, \gamma_{j,y,2}, \cdots, \gamma_{j,y,k})$, with $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\gamma})$.

As before, a joint model $\boldsymbol{\pi}(\boldsymbol{y} \mid x, \boldsymbol{\theta}) = \Pr(\mathbf{Y} = \boldsymbol{y} \mid x, \boldsymbol{\theta})$ may be obtained from the marginals by using a 1-parameter copula. The number of model parameters, $p = k(m_T + m_E) + 1$, may seem large for a dose-finding design model. For example, if $(k, m_E, m_T) = (5,3,3)$, then $p = 31$. In fact, the method based on this model performs quite well in practice. Implementing MCMC for computing posteriors is not problematic, and the design has good properties well across a wide set of dose-outcome scenarios.

A simpler, more conventional alternative marginal model for ordinal Y_j is given by $g(\bar{\pi}_{j,y,x}) = \alpha_{j,y} + \beta_j x$, where $\alpha_{j,y}$ decreases in y for each k. This is the proportional odds model (McCullagh, 1980) if g is the logit link. For example, in the case of four levels for each outcome, $m_T = m_E = 3$ and $p = 9$.

4.2 Computing Expected Utilities

Denoting the elicited utility of outcome $\mathbf{y}$ by $U(\mathbf{y})$, the *mean utility of dose x given $\boldsymbol{\theta}$* is

$$u(x, \boldsymbol{\theta}) \;=\; \sum_{y_1=0}^{m_1} \sum_{y_2=0}^{m_2} U(\boldsymbol{y})\, \boldsymbol{\pi}(\boldsymbol{y} \mid x, \boldsymbol{\theta}).$$

Since $u(x, \boldsymbol{\theta})$ is a theoretical quantity depending on the parameter vector $\boldsymbol{\theta}$, to obtain a statistic that can be used for decision-making the posterior mean of $u(x, \boldsymbol{\theta})$ is computed for each x. The *posterior mean utility of dose x given data $\mathcal{D}_n$* is

$$\xi(x, \mathcal{D}_n) \;=\; \mathrm{E}_{\boldsymbol{\theta}}\{u(x, \boldsymbol{\theta}) \mid \mathcal{D}_n\} \;=\; \sum_{y_1=0}^{m_1} \sum_{y_2=0}^{m_2} U(\boldsymbol{y}) \int_{\boldsymbol{\theta}} \boldsymbol{\pi}(\boldsymbol{y} \mid x, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathcal{D}_n) d\boldsymbol{\theta}. \tag{4.4}$$

This combines the physician's utilities and the data in terms of a single numerical criterion quantifying the desirability of x. We denote the dose that maximizes $\xi(x, \mathcal{D}_n)$ by x_n^{opt}.

4.3 Safety

To control the risk of toxicity, let y^* be the level of toxicity considered to be unacceptable elicited from the physician, and π_1^* a fixed limit on $\bar{\pi}_{1,y^*,x}$. *A dose x is unacceptably toxic* if

$$\Pr(\bar{\pi}_{1,y^*,x} > \pi_1^* \mid \mathcal{D}_n) > p_U, \tag{4.5}$$

where p_U is an upper probability cut-off, such as .90. *The set of acceptably safe doses*, $\mathcal{A}_n(\bar{\pi})$, is defined to be all $x \in \mathcal{X}$ for which (4.5) is not the case.

An additional safety constraint is that an untried dose may not be skipped when escalating. Essentially, this limits extrapolation based on the model to one dose level above the highest dose at which patients have been treated.

4.4 Adaptive Randomization

A common practical problem when selecting doses adaptively is that, in some cases, little or no information may be obtained for the dose that actually has the highest true mean utility. The algorithm that simply assigns x_n^{opt} is an example of a "greedy algorithm" and it may get stuck, repeatedly assigning a suboptimal dose and failing to find the true optimal dose. This property of greedy sequential decision procedures is well-known. A solution to this problem is to randomly assign some patients to nearly optimal treatments, which distributes patients more evenly and thus allows the one to learn more about the treatment space, with a resulting improvement in the procedure's reliability.

Table 4: Operating characteristics of the malignant melanoma B-RAF inhibitor design

	x = 1	x = 2	x = 3	None
	Scenario 1			
True $(\pi_E, \pi_T), u$	(.45, .05), 48.9	(.60, .10), 60.3	(.75, .15), 71.2	
% Selected	5	18	74	3
# Treated	8.4	10.2	16.7	
	Scenario 2			
True $(\pi_E, \pi_T), u$	(.45, .05), 48.9	(.65, .08), 65.4	(.67, .25), 56.9	
% Selected	9	69	18	4
# Treated	9.9	15.3	9.9	
	Scenario 3			
True $(\pi_E, \pi_T), u$	(.50, .45), 40.9	(.55, .50), 42.7	(.60, .55), 44.7	
% Selected	4	1	0	95
# Treated	9.6	3.7	1.8	

To do this ethically in dose-finding trial, the following adaptive randomization (AR) procedure may be applied.

Given a sequence $\boldsymbol{\delta} = \{\delta_n,\ n = 1, \cdots, N_{max}\}$ of non-increasing utility differences, *the set of* δ_n*-optimal doses* is

$$\mathcal{A}_n(\bar{\pi}, \delta) = \{x \in \mathcal{A}_n(\bar{\pi}) : \xi(x, \mathcal{D}_n) \geq \xi(x_n^{opt}, \mathcal{D}_n) - \delta_n\}. \tag{4.6}$$

In words, $\mathcal{A}_n(\bar{\pi}, \delta)$ is the set of safe doses having posterior mean utility within δ_n of the maximum. The sequence $\boldsymbol{\delta}$ is chosen to be non-increasing with n to accommodate the decreasing variability in the posterior mean utilities. Patients are randomized among the doses in $\mathcal{A}_n(\bar{\pi}, \delta)$. The AR probabilities can be chosen in many ways, but a simple and reliable approach is to weight the doses in $\mathcal{A}_n(\bar{\pi}, \delta)$ equally. The use of AR is motivated by both ethical considerations and the fact that the posteriors of the utilities $\{u(x, \boldsymbol{\theta}), x \in \mathcal{A}_n(\bar{\pi})\}$ may be quite disperse. Extensive simulations (Thall and Nguyen, 2012) show that, perhaps counter-intuitively, it often is more ethical to treat some patients at suboptimal doses having $\xi(x, \mathcal{D}_n)$ near $\xi(x_n^{opt}, \mathcal{D}_n)$. This is because, on average, application of AR causes more patients in the trial to be treated at doses having higher true utilities.

For trial conduct, the first cohort is treated at a starting dose chosen by the physician. For all subsequent cohorts, patients are randomized among the acceptable doses using the updated AR probabilities given $\mathcal{D}_n$. If $\mathcal{A}_n(\bar{\pi}, \delta)$ is empty then the trial is stopped and no dose is chosen. A rule superseding the above is that no untried dose may be skipped when escalating. At the end of the trial, if $\mathcal{A}_n(\bar{\pi}, \delta)$ is not empty, $x_{N_{max}}^{opt}$ is selected.

4.5 Illustration: A Dose-Finding Trial in Malignant Melanoma

It is believed that mutations of the B-RAF gene may cause cancer. A phase I/II trial was designed to optimize dose of a new B-RAF inhibitor, given with tumor infiltrating lymphocytes, to treat malignant melanoma. Three doses of the B-RAF inhibitor were considered, { 320, 640, 960} mg given twice daily. Toxicity was defined as any grade 3 or 4 non-hematologic toxicity occurring within four weeks from the start of treatment and not resolved therapeutically within two weeks. Efficacy was defined as immunological response. The outcome thus was bivariate binary, with indicators Y_T and Y_E. The method was applied with cohorts of size three, starting at x = 1, choosing x_n^{opt} for each cohort until 9 patients were treated, with AR was applied thereafter, up to a maximum of 36 patients.

In the bivariate binary outcome case, $m_1 = m_2 = 1$, and $p = 2k + 1$. The only marginal probabilities are $\pi_{E,x}$ and $\pi_{T,x}$ for x = 1,2,3, and $\boldsymbol{\theta}_j = (\theta_{j,1}, \cdots, \theta_{j,k})$, for $j = 1, 2$. Here, p =7 since k = 3. The elicited prior means were $\mathrm{E}(\pi_{E,1}, \pi_{E,2}, \pi_{E,3})$ = (.45, .65, .75) and $\mathrm{E}(\pi_{T,1}, \pi_{T,2}, \pi_{T,3})$ = (.05, .10, .15). Given the two best and worst elementary event utilities fixed at $U_{1,0}$ = 100 for (Y_E, Y_T) = (1,0) and $U_{0,1}$ = 0 for (Y_E, Y_T) = (0,1), the elicited utilities of the remaining two intermediate elementary events were $U_{1,1}$ = 50 and $U_{0,0}$ = 10. For safety monitoring, a dose x was considered unacceptably toxic if $\Pr(\pi_{T,x} > .25|\mathcal{D}_n) > .80$, and unacceptably inefficacious if $\Pr(\pi_{E,x} < .35|\mathcal{D}_n) > .80$. The operating characteristics of this design are summarized in Table 4. The simulations show that the method does a good job of selecting doses having the highest utilities with high probabilities (Scenarios 1 and 2), and stopping with high probability when no dose is safe (Scenario 3).

4.6 Acknowledgement

This research was partially supported by NCI grant RO1-CA-83932. The author is grateful to a referee for very detailed and constructive comments on an earlier draft.

References

[1] Atkinson, A.C., Donev, A.N. and Tobias, R. (2007). Optimum Experimental Designs, with SAS. Oxford University Press, Oxford, UK.

[2] Bretz, F., Pinheiro, J.C. and Branson, M. (2005) Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*, **61**, 738–748.

[3] Bryant, J. and Day, R. (1995) Incorporating toxicity considerations into the design of two-stage phase II clinical trials. *Biometrics* **51**, 1372–1383.

[4] Chevret, S. (2006) *Statistical Methods for Dose-Finding Experiments*. John Wiley and Sons, West Sussex, UK.

[5] Cheung, Y-K. (2011) *Dose Finding by the Continual Reassessment Method*. Chapman and Hall/CRC Press, New York.

[6] Conaway, M.R. and Petroni, G.R.(1995) Bivariate sequential designs for phase II trials. *Biometrics* **51**, 656–664.

[7] Dougherty, T.B., Porsche, V.H. and Thall, P.F. (2000) Maximum tolerated dose of nalmefene in patients receiving epidural fentanyl and dilute bupivaccine for postoperative analgesia. *Anesthesiology*, **92** 1010–1016.

[8] Fleming, T.R. (1982) One-sample multiple testing procedure for phase II clinical trials. *Biometrics* **38**, 143–51.

[9] Houede, N., Thall, P.F., Nguyen, H., Paoletti, X. and Kramar, A. (2010) Utility-based optimization of combination therapy using ordinal toxicity and efficacy in phase I/II trials. *Biometrics*. **66** 532–540.

[10] Le Tourneau, C., J. Lee, J.J. and Siu, L.L. (2009) Dose escalation methods in phase I cancer clinical trials. *J. National Cancer Institute*, **101**, 708–720.

[11] Morita, S., Thall, P.F. and Mueller, P. (2010) Evaluating the impact of prior assumptions in Bayesian biostatistics. *Statistics in Biosciences* **2** 1–17.

[12] Nelsen, R.B. *An Introduction to Copulas. Lecture Notes in Statistics, 139*. New York, Springer-Verlag, 1999.

[13] O'Quigley, J., Pepe, M. and Fisher, L. (1990) Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* **46**, 33–48.

[14] Rubinstein, L.V., Korn, E.L., Freidlin, B., Hunsberger, S., Ivy, P. and Smith, M. (2005) Design issues of randomized phase II trials and a proposal for phase II screening trials. *J. Clinical Oncology*, **23**, 7199–7206.

[15] Simon, R. (1989) Optimal two-stage designs for phase II clinical trials. *Controlled Clin Trials* **10**, 1–10.

[16] Simon, R., Wittes, R.E. and Ellenberg, S.S. (1985) Randomized phase II clinical trials. *Cancer Treatment Reports* **69**, 1375–1381.

[17] Taylor, J.M.G., Braun, T.M. and Li, Z. (2006) Comparing an experimental agent to a standard agent: relative merits of a one-arm or randomized two-arm phase II design. *Clinical Trials* **3**, 335–338.

[18] Thall, P.F. and Cook, J.D. (2004). Dose-finding based on efficacy-toxicity trade-offs. *Biometrics* **60**, 684–693.

[19] Thall, P.F., Cook, J.D., and Estey, E.H. (2006). Adaptive dose selection using efficacy-toxicity trade-offs: illustrations and practical considerations. *Journal of Biopharmaceutical Statistics* **16**, 623–638.

[20] Thall, P.F. and Nguyen, H.Q. (2010) Covariate-adjusted adaptive dose-finding in early phase clinical trials. In: Chow SC, ed. *Encyclopedia of Biopharmaceutical Statistics. 3rd Edition*. London: Informa Healthcare Ltd. **1**, 369–379.

[21] Thall, P.F. and Nguyen, H.Q. (2012) Adaptive randomization to improve utility-based dose-finding with bivariate ordinal outcomes. *Journal of Biopharmaceutical Statistics*, **22**, 785–801.

[22] Thall, P.F., Nguyen, H.Q., and Estey, E.H. (2008). Patient-specific dose-finding based on bivariate outcomes and covariates. *Biometrics* **64**, 1126–1136.

[23] Thall, P.F. and Russell, K.T. (1998). A strategy for dose finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Biometrics* **54**, 251–264.

[24] Thall, P.F. and Simon, R. (1994) Practical Bayesian guidelines for phase IIB clinical trials. *Biometrics* **50**, 337–349.

[25] Thall, P.F., Simon, R. and Estey, E.H. (1995) Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. *Statistics in Medicine* **14**, 357–379.

[26] Thall, P.F., Szabo, A., Nguyen, H.Q., Amlie-Lefond, C.M. and Zaidat, O.O. (2011) Optimizing the concentration and bolus of a drug delivered by continuous infusion. *Biometrics* **67**, 1638–1646.

[27] Whelan, H.T., Cook, J.D., Amlie-Lefond, C.M., Hovinga, C.A., Chan, A.K., Ichord, R.N., deVeber, G.A., and Thall, P.F. (2008) Practical model-based dose-finding in early phase clinical trials: Optimizing tissue plasminogen activator dose for treatment of ischemic stroke in children. *Stroke* **39**, 2627–2636.

[28] Zohar, S. and Chevret, S. (2007). Recent developments in adaptive designs for phase I/II dose-finding studies. *Journal of Biopharmaceutical Statistics* **17**, 1071–1083.