BIOMETRICS 57, 914–921
September 2001

Dose-Finding Based on Feasibility and Toxicity in T-Cell Infusion Trials

Peter F. Thall,[1,*] Hsi-Guang Sung,[1] and Aniruddha Choudhury[2]

[1]Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Box 447, 1515 Holcombe Boulevard, Houston, Texas 77030, U.S.A.
[2]Department of Oncology Pathology, Cancer Centrum Karolinska, R8.01 Stockholm, Sweden 17176
*email: rex@mdanderson.org

SUMMARY. A new modality for treatment of cancer involves the *ex vivo* growth of cancer-specific T-cells for subsequent infusion into the patient. The therapeutic aim is selective destruction of cancer cells by the activated infused cells. An important problem in the early phase of developing such a treatment is to determine a maximal tolerated dose (MTD) for use in a subsequent phase II clinical trial. Dose may be quantified by the number of cells infused per unit body weight, and determination of an MTD may be based on the probability of infusional toxicity as a function of dose. As in a phase I trial of a new chemotherapeutic agent, this may be done by treating successive cohorts of patients at different dose levels, with each new level chosen adaptively based on the toxicity data of the patients previously treated. Such a dose-finding strategy is inadequate in T-cell infusion trials because the number of cells grown *ex vivo* for a given patient may be insufficient for infusing the patient at the current targeted dose. To address this problem, we propose an algorithm for trial conduct that determines a feasible MTD based on the probabilities of both infusibility and toxicity as functions of dose. The method is illustrated by application to a dendritic cell activated lymphocyte infusion trial in the treatment of acute leukemia. A simulation study indicates that the proposed methodology is both safe and reliable.

KEY WORDS: Adoptive immunotherapy; Clinical trial design; Continual reassessment method; Dose finding; Leukemia; Maximal tolerated dose; Phase I trial; T-cell infusion.

1. Introduction

In recent years, numerous studies in animals and humans have provided extensive data that T-cells, which are specific types of white blood cell, may be utilized for the treatment of cancer (Cheever, Greenberg, and Fefer, 1981; Porter et al., 1994). This is because, when properly stimulated, T-cells kill cancer cells. Various versions of this new therapeutic modality have been referred to as adoptive immunotherapy, donor lymphocyte infusion, or T-cell immunotherapy. Initial trials have reported significant therapeutic responses in humans. These trials, taken together with rapidly emerging technology whereby specific types of T-cells are grown *ex vivo,* i.e., outside the patient's body, have motivated a wide variety of T-cell immunotherapy trials. The basic therapeutic approach uses cells taken either from the patient (autologous cells) or from a donor (allogeneic cells) as the starting material. These cells are expanded *ex vivo* and then infused into the patient, possibly after an initial course of chemotherapy. Because currently these are early-phase trials, there has been a heavy reliance on animal data and anecdotal experience to determine cell dose, which is the number of cells infused, either overall or per unit of the patient's body weight. Determining a safe and efficacious dose in a particular clinical setting is problematic, however, due to the fact that severe, sometimes fatal, infusional toxicities may occur (Freedman et al., 1994; Figlin et al., 1997; Plautz et al., 1998). Such toxicities may take the form of fever, myalgia, nausea, and adverse effects in the liver, kidneys, or blood. Unfortunately, to date, little attention has been paid to dose finding in T-cell infusion trials.

Although T-cell infusional toxicities are often qualitatively different from adverse events typically associated with chemotherapy, infusional toxicities are more likely with higher doses of cells infused (Alyea et al., 1998). This fact, and the implicit assumption that a higher dose may provide more anticancer effect make the problem of determining an acceptable dose analogous to that in conventional chemotherapy trials. Due to the monotonicity of the dose-toxicity curve, established dose-finding methods appropriate for phase I chemotherapy trials, such as the continual reassessment method (CRM; O'Quigley, Pepe, and Fisher, 1990) should also provide a basis for dose-finding in T-cell infusion trials. The CRM has been shown in several simulation studies to be superior to conventional 3+3 algorithms for dose finding in phase I (Goodman, Zahurak, and Piantadosi, 1995; Møller, 1995; Ahn, 1998). The CRM is more likely than 3+3 algorithms to select as the maximal tolerated dose that dose having toxicity probability closest to a prespecified target probability, and on average, the CRM treats more patients at or near the MTD. For these reasons

and because computer software required for trial design and conduct has become freely available, the CRM is now being used increasingly in phase I trials (Piantadosi, Fisher, and Grossman, 1998; Thall et al., 1999; Dougherty, Porsche, and Thall, 1999). Thus, to determine an MTD, it might appear that if "number of cells infused per kilogram body weight" is substituted for "dose of chemotherapy," then the CRM can be applied.

An additional problem may arise, however, due to the fact that the cells to be infused into the patient must be grown *ex vivo*. For a given patient with a given targeted dose in the context of a given dose-finding algorithm, the number of cells grown may be below the targeted infusion dose for the patient. This raises the questions of how to proceed therapeutically for such a patient, how to evaluate the feasibility of the approach while also evaluating toxicity, and how to use the available infusibility and toxicity data on that basis to sequentially select doses for patient cohorts during the trial. Thus, the feasibility of the therapeutic strategy, in addition to the more usual considerations involving toxicity, must be considered in the design and conduct of the trial. The present article is motivated by this problem. We propose an algorithm for trial conduct based on the scientific and practical goal of determining a feasible maximal tolerated dose (FMTD).

The remainder of the article is organized as follows. The T-cell infusion trial that motivated this research is described in Section 2. In Section 3, we establish probability models and decision criteria that account for both toxicity and feasibility. We present the dose-finding algorithm and trial design in Section 4 and describe a simulation study of the dendritic cell (DC) activated lymphocytes (AL) infusion trial in Section 5. We close with a discussion in Section 6.

2. A T-Cell Infusion Trial in Acute Leukemia

Patients with acute myelogenous leukemia who have achieved a remission with chemotherapy but subsequently relapsed in less than 2 years have a very poor prognosis. Conventional chemotherapy salvage treatments are unlikely to achieve a second remission in these patients, and median survival is less than 6 months. Patients with chronic myelogenous leukemia who have entered an advanced stage of the disease known as blast crisis have a similarly poor prognosis. Consequently, the need to develop and clinically evaluate innovative treatments for these patients is quite pressing.

One relatively new type of T-cell immunotherapy involves the use of dendritic cell activated lymphocytes (DC-AL). This approach begins with harvesting of cells from the patient's circulating blood or, if necessary, from the bone marrow. *Ex vivo,* two types of cells are grown, the first consisting of mostly leukemic cells and the second of mostly T-lymphocytes. A combination of growth factors is used to stimulate the leukemia cells such that they differentiate into DCs, and this cell culture is then irradiated to kill any remaining growing leukemia cells. The two cell types are then cocultured, with the goal that the DCs will stimulate leukemia-specific lymphocytes and thereby increase their ability to recognize and kill leukemia cells. The patient is given preliminary chemotherapy, followed by infusion of the DC-AL mixture. The rationale for this therapeutic strategy is based on laboratory data showing that DCs can be grown in culture from chronic myelogenous leukemia cells (Choudhury et al., 1997) and from acute myelogenous leukemia cells (Choudhury et al., 1999) and on the clinical trials of similar strategies in other cancers noted earlier. This therapeutic strategy is very different from conventional bone marrow transplantation, where intensive chemotherapy or radiation is used to ablate the patient's bone marrow prior to cell infusion, with the aim that the infused cells will repopulate the patient's marrow.

The DC-AL infusion trial includes both relapsed acute myelogenous leukemia patients and chronic myelogenous leukemia patients in blast crisis. The primary scientific goal is to determine an MTD of cells to infuse. This MTD will be used in a future phase II clinical trial of this new therapeutic modality, which will be conducted once the phase I trial is complete. As always in trials of novel treatments involving adverse events, due to ethical considerations, dose finding must be done sequentially, with the dose for each successive cohort of patients determined based on the doses and clinical outcomes of patients treated previously in the trial. For the purpose of dose finding in the DC-AL infusion trial, toxicity is defined to include any grade 3 or 4 (severe) nonhematologic toxicity. Anticipated nonhematologic toxicities are of two types, either those related to the cell infusion or autoimmune phenomena related to the effects of the activated lymphocytes, which may attack nonleukemia cells in the patient. Infusional toxicities include fever, shortness of breath, and/or hypotension. Autoimmune toxicities include skin rash, noninfectious pneumonia, kidney inflammation, and arthritis. Thus, the risk of infusional and autoimmune toxicities is a serious problem. Because these toxicities are biological consequences of the infusion and subsequent lymphocyte activity in the patient, necessarily, the probability of toxicity is an increasing function of dose.

3. Probability Models

The decision criteria underlying the algorithm for trial conduct are based on Bayesian models for the probabilities of infusibility and toxicity. The Bayesian paradigm provides a natural probability structure for repeatedly incorporating new information as it accumulates during the trial and making decisions sequentially on that basis. We formulate these models as parsimoniously as is reasonable, both to provide generality and to facilitate numerical computations. The latter is a practical consequence of our requirement that the design be parameterized so that it has good operating characteristics (OCs) since this in turn requires simulation of the trial within a reasonable time frame during the process of developing a design. In practice, the OCs may be used as a basis for calibrating the design parameters. Because trial conduct will rely on decision criteria that require updating the posterior distribution of the model parameters as the data from successive patients become available, each simulation of the trial requires many such posterior computations. Here, the OCs will consist of dose selection, early-termination probabilities, and the average numbers of patients treated and experiencing toxicity at each dose. We compute these values under each of several dose–outcome scenarios. In general, we consider a design's OCs to be good only if both the physician(s) organizing the trial and the statistician providing the design consider them so since these individuals are responsible for trial conduct.

3.1 *Infusibility and Feasibility*

Let L denote the number of DC-ALs grown *ex vivo* for a given patient and denote the number of dose levels by k. Since the relevant dose levels are determined by the particular numerical cutoffs $d_1 < d_2 < \cdots < d_k$ for L used in the trial, the random variable that matters is not L *per se* but rather the maximum dose level at which the patient may be infused, given formally by $Y = \Sigma_{j=1}^{k} jI(d_j \le L < d_{j+1})$, i.e., $(Y = j) = (d_j \le L < d_{j+1})$ is the event that enough cells have been grown *ex vivo* to infuse the patient at any targeted dose level up to j but not above j and $[Y = 0]$ is the event that there are not enough cells to infuse the patient even at the lowest dose level. A patient for whom Y is below the targeted dose level still will be infused at Y since this is the patient's best hope therapeutically. Patients for whom $Y = 0$ must be treated under some alternative protocol, however, since d_1 is considered to be the lowest dose having potential therapeutic benefit. Thus, among n patients who are accrued and have cells extracted, the number of patients who are infused at some level and evaluated for toxicity is $\Sigma_{i=1}^{n} I(Y_i \ge d_1)$, which may be smaller than n.

Because the process of generating dendritic cells *ex vivo* precedes infusion and hence the phenomenon of the patient experiencing toxicity, we evaluate feasibility based on the marginal probability distribution of Y. As we will show in Section 3, however, toxicity and Y are not independent. We will present two different marginal probability models for Y. The choice of a model for use in a particular trial will depend on practical considerations, including numerical computation and whether the model yields a design with good OCs. In general, we will denote the probability of infusibility at dose level j by $\theta_j = \Pr(Y \ge j)$ and denote $\pi_j = \Pr(Y = j) = \theta_j - \theta_{j+1}$, where $\theta_{k+1} = 0$ and $\theta_0 = 1$.

Model 1 is obtained by determining k fixed probabilities $q_1 > q_2 > \cdots > q_k$ and defining

$$\theta_j = q_j^{\exp(\beta)}, \qquad j = 1, \ldots, k, \tag{1}$$

where *a priori* $\beta \sim \mathrm{N}(0, \sigma_\beta^2)$. Since q_j is the value of θ_j for $\beta = 0$, the q_j's may be obtained by first eliciting prior average infusibility probabilities from the physician and then, if necessary, calibrating the elicited values on the basis of simulation results to obtain a design with good overall OCs. Thus, the q_j's are fixed parameters that determine the marginal distribution of Y along with the random parameter, β. Given the q_j's, model 1 is characterized by β, with $\pi_j = q_j^{\exp(\beta)} - {q_{j+1}}^{\exp(\beta)}$, denoting $q_0 = 1$ and $q_{k+1} = 0$. We assume a reasonably uninformative Gaussian prior for β by taking $\sigma_\beta^2 = 1$ to reflect considerable prior uncertainty on the part of the physicians and to ensure that, early in the trial, the data are the main ingredient in determining the successive posteriors and the decisions based on them.

If one does not wish to rely on the fixed values $q_1, \ldots, q_k$ of model 1, which do not change once the model is established, an alternative is model 2. This is obtained by assuming that $\boldsymbol{\pi} = (\pi_0, \ldots, \pi_{k-1})$ follows a Dirichlet prior with parameters $\mathbf{a} = (a_0, \ldots, a_k)$, denoted $\boldsymbol{\pi} \sim \mathrm{Dir}(\mathbf{a})$. Since $\mathrm{E}(\pi_j) = \pi_j^o = a_j/a_+$, where $a_+ = a_0 + \cdots + a_k$ quantifies the amount of information in the prior, the a_j's may be determined by eliciting prior mean infusibility probabilities $\{\pi_j^o,\ j = 0, \ldots, k-1\}$ from the physicians and then calibrating a_+ based on the degree of prior knowledge about the π_j's. Thus, π_j^o under model 2 corresponds to $q_j - q_{j+1}$ under model 1. In the DC-AL infusion trial, we elicited $(q_0, \ldots, q_5) = (1, 0.975, 0.95, 0.90, 0.75, 0.50)$, equivalently $(\pi_0^o, \ldots, \pi_5^o) = (0.025, 0.025, 0.05, 0.15, 0.25, 0.50)$. At any given point in the trial, denote the number of patients for whom $Y = j$ by X_j and $\mathbf{X} = (X_0, \ldots, X_k)$. Since $[\mathbf{X} \mid \boldsymbol{\pi}]$ is multinomial in $\boldsymbol{\pi}$ and $n = X_0 + \cdots + X_k$, the posterior is $[\boldsymbol{\pi} \mid \mathbf{X}] \sim \mathrm{Dir}(\mathbf{a}+\mathbf{X})$. While model 2 may appear to be more complex, its implementation requires the same information as that required by model 1. Since the π_j^o's replace the q_j's, and the Dirichlet as prior is conjugate to the multinomial, the numerical computations under model 2 are no more complex than those for model 1. The likelihood based on the infusibility data is

$$\mathcal{L}_{\mathrm{infuse}}(\boldsymbol{\pi}) = \prod_{j=0}^{k} {\pi_j}^{X_j} \tag{2}$$

under model 2. Under model 1, the parameter vector is simply β, and ${q_j}^{\exp(\beta)} - {q_{j+1}}^{\exp(\beta)}$ replaces π_j in (2).

Our dose-finding algorithm uses the following feasibility criterion, defined in terms of the infusibility probabilities, in conjunction with the CRM's more usual toxicity criterion. Let θ^* be a fixed minimum required probability of infusibility that is specified by the physicians conducting the trial and let $p_{\mathrm{U,F}}$ be an upper probability cutoff, taking on typical values in the range 0.90–0.99.

DEFINITION: *Dose level j is not feasible if*

$$\Pr(\theta_j < \theta^* \mid \mathrm{data}) > p_{\mathrm{U,F}}. \tag{3}$$

Thus, a dose level is considered feasible only if it is not very likely that its infusibility probability is below the minimum level specified by the physicians. Since a given dose level being not feasible implies that no higher level is feasible, the set of feasible doses at any point in the trial is of the form $\{1, \ldots, j\}$ for some $j \le k$. Under model 1, $[\theta_j < \theta^*] = [\beta > \log\{\log(\theta^*)/\log(q_j)\}]$, so the one-dimensional numerical integration necessary to determine the set of feasible doses is straightforward. The feasibility criterion also may easily be computed under model 2 using standard software by exploiting the fact that $\theta_j = \pi_j + \cdots \pi_k$ has a beta posterior with parameters $\Sigma_{r=j}^{k} (a_r + X_r)$ and $\Sigma_{l=0}^{j-1} (a_l + X_l)$ for each $j = 1, \ldots, k$.

3.2 *Toxicity*

To implement the portion of the design based on toxicity, we use the continual reassessment method (CRM) with the following exponential probability model, which is formally equivalent to the model originally proposed by O'Quigley et al. (1990). The model specifies the probability of toxicity as an increasing function of dose level. Let T be the binary indicator of toxicity, fix the k probabilities $p_1 < p_2 < \cdots < p_k$, and define $\Pr(T = 1 \mid \text{infused at } d_j) = p_j^{\exp(\alpha)}$, where *a priori* $\alpha \sim \mathrm{N}(0, \sigma_\alpha^2)$. We chose this prior for numerical convenience and to reflect little prior knowledge about toxicity, using $\sigma_\alpha^2 = 1.34$, as suggested by O'Quigley and Shen (1996). Although this CRM model is similar to model 1 for infusibility, the

fundamental difference is that $\Pr(T = 1 \mid d_j)$ is assumed to be increasing in d_j for biological reasons while $\Pr(Y \geq j)$ must be decreasing in j to ensure that $\{\theta_j\}$ is a probability distribution. The probability distribution of T is specified as a function of the dose level at which the patient is actually infused because this is what is relevant for dose finding. However, because d_j depends on both the number of cells grown *ex vivo*, Y, and the algorithm for trial conduct, the distribution of T depends on Y. This will be explained later in Section 3. In practice, when designing a dose-finding trial based on toxicity alone using the CRM, we choose the p_j's by first eliciting values from the physician and then adjusting them, based on preliminary simulations, to obtain a design with good OCs. For each successive patient cohort, the CRM selects the dose that has mean posterior toxicity probability, $\mathrm{E}(p_j^{\exp(\alpha)} \mid \text{data})$, closest to the target value p^* specified by the physician. In the DC-AL infusion trial, $p^* = 0.30$, $k = 5$, $(p_1, \ldots, p_5) = (0.05, 0.10, 0.30, 0.50, 0.60)$. We also imposed the safety modification that no dose level may be skipped when escalating. In a more usual dose-finding trial using the CRM and based on toxicity alone, the maximal tolerated dose (MTD) is simply the dose selected by the CRM at the end of the trial. In the present setting, the CRM criterion is used in conjunction with a criterion pertaining to feasibility, described below. Under this model, the p_j's are fixed while α is random; hence, the toxicity data are used to update only the distribution of α. The likelihood for the toxicity data of n infused patients is

$$\mathcal{L}_{\mathrm{tox}}(\alpha) = \prod_{i=1}^{n} \left\{ p_{j(i)}^{\exp(\alpha)} \right\}^{T_i} \left\{ 1 - p_{j(i)}^{\exp(\alpha)} \right\}^{1-T_i}, \qquad (4)$$

where $j(i)$ denotes the dose level at which patient i is infused. To ensure a safe trial, in addition to the usual CRM criterion, we also require the following definition. Let $p_{\mathrm{U,T}}$ be an upper probability cutoff analogous to $p_{\mathrm{U,F}}$.

DEFINITION: *Dose level j is unacceptably toxic if*

$$\Pr\left(p_j^{\exp(\alpha)} > p^* \mid \text{data} \right) > p_{\mathrm{U,T}}. \qquad (5)$$

If a given dose level is unacceptably toxic, then so are all higher dose levels. The practical implication is that the trial is terminated early if the lowest dose level is unacceptably toxic. The criterion (5) was used by Thall and Russell (1998) to define a dose-finding algorithm under a more complex model accommodating both toxicity and response. It is similar in aim to the so-called overdose control rule, which sets limits on both the dose and the toxicity probability, proposed by Babb, Rogatko, and Zacks (1998). We will illustrate the effect of the additional safety rule (5) in the context of the DC-AL infusion trial in Section 4.

4. Dose-Finding Algorithm

At any given point in the trial, denote the current targeted dose by j^* and denote the dose recommended by the CRM by j_{CRM}. The set of feasible doses must be of the form $\{1, \ldots, j\}$, with $\{j+1, \ldots, k\}$ unacceptable. The case where $j_{\mathrm{CRM}} \leq j$ is nonproblematic since j_{CRM} is feasible; hence, we simply set $j^* = j_{\mathrm{CRM}}$. If $j_{\mathrm{CRM}} > j$, the case in which the dose recommended by the CRM is not feasible, then we set $j^* = j$, the feasible dose closest to the dose recommended by the CRM. Thus, in general, $j^* \leq j_{\mathrm{CRM}}$. Let c denote the number of patients in each cohort evaluated for toxicity at each successive target dose level. Let N be the total number of patients to be evaluated for toxicity and let M be the maximum number of patients that can have cells extracted and evaluated for infusibility. The parameter N is the sample size corresponding to a more common dose-finding trial using the CRM and based on toxicity alone. Since some patients may have cells extracted but may not be infused due to the fact that an insufficient number of cells were grown to infuse the patient even at the lowest dose level ($Y = 0$), M should be specified as an absolute upper limit due to resource limitations, with $N \leq M$. The trial is conducted as follows:

1. Begin the trial with target dose level j^* having prior mean $\mathrm{E}(p_j^{\exp(\alpha)})$ closest to p^*.
2. At each j^*, in order of accrual, extract and process cells from each patient and determine Y. Once T is evaluated for all previous patients, determine j_{CRM} and j^*. If $Y \geq j_{\mathrm{CRM}}$, then infuse the patient at j_{CRM} and evaluate T. If $Y < j_{\mathrm{CRM}}$, then infuse the patient at Y and evaluate T, unless $Y = 0$. Continue in this manner until, based on the most recent posterior information, either (a) j^* is determined to be not feasible or (b) c patients have been infused and evaluated for toxicity at j^*.
3. Once either criterion 2a or 2b has been met at the current j^*, the next targeted dose level will be the feasible dose level having mean posterior toxicity probability closest to θ^*, subject to the constraint that no untried dose level may be skipped when escalating.
4. The trial ends when either (a) N patients have been infused and evaluated for toxicity or (b) M patients have had cells extracted, regardless of infusibility. In either case, the FMTD is the highest feasible dose level $\leq j_{\mathrm{CRM}}$ based on the posterior from the final data.
5. If it is determined that the lowest dose level is either not feasible or unacceptably toxic, then the trial is terminated.

Rule 2 formalizes the practical goal of targeting only feasible doses. If $j^* < j_{\mathrm{CRM}}$ and $j^* < Y$ for a particular patient, however, then it is ethically and scientifically appropriate to take advantage of the fact that more DCs than targeted were grown and infuse the patient at $\min\{Y, j_{\mathrm{CRM}}\}$, the highest possible dose level $\leq j_{\mathrm{CRM}}$. Thus, determining the set of feasible doses allows the targeted level to be below j_{CRM}. Under rule 5, the trial is terminated early if no dose level is feasible. Rule 3 simply says that the usual CRM criterion will be used to select the next targeted dose level, with the important constraint that only feasible dose levels may be considered. Rule 5 also ensures that, aside from feasibility, if the lowest dose level is excessively toxic, the trial will be terminated early on ethical grounds. If the trial is terminated early due to rule 5, then no FMTD is determined.

Although the distribution of T that is relevant for dose finding is conditional on the dose level at which the patient is infused rather than the number of cells grown *ex vivo*, T and Y are not independent. In fact, the distribution of T depends on Y in a manner determined by both the marginal distribution of Y and the dose-finding algorithm. Specifically,

Table 1

Operating characteristics of the DC-AL infusion trial design under models 1 *or* 2. p_{tox} = Pr(toxicity) *and* p_{inf} = Pr(infusibility). *Under each scenario, numbers in blocks corresponding to either the optimal dose or the correct decision of choosing no dose are given in boldface. The column labeled None contains early termination percentages. Each entry is the simulated mean and its standard deviation.*

Scenario		Dose Level					
		1	2	3	4	5	None
1	$(p_{\text{tox}}, p_{\text{inf}})$	(0.10, 0.99)	**(0.30, 0.95)**	(0.50, 0.90)	(0.70, 0.75)	(0.80, 50)	—
	No. infused	1.7 (0.02)	**10.4 (0.08)**	9.7 (0.08)	0.7 (0.02)	0.0 (0.00)	—
	No. of toxicities	0.2 (0.00)	**3.1 (0.03)**	4.8 (0.03)	0.5 (0.01)	0.0 (0.00)	—
	% FMTD	2.9 (0.17)	**60.7 (0.5)**	27.8 (0.45)	0.1 (0.03)	0.0 (0.00)	8.5 (0.28)
2	$(p_{\text{tox}}, p_{\text{inf}})$	(0.10, 0.90)	**(0.30, 0.75)**	(0.50, 0.50)	(0.70, 0.25)	(0.80, 0.05)	—
	No. infused	4.6 (0.03)	**13.0 (0.06)**	4.0 (0.04)	0.2 (0.01)	0.0 (0.00)	—
	No. of toxicities	0.5 (0.01)	**3.9 (0.02)**	2.0 (0.02)	0.1 (0.01)	0.0 (0.00)	—
	% FMTD	3.8 (0.19)	**74.5 (0.44)**	10.3 (0.30)	0.1 (0.03)	0.0 (0.01)	11.3 (0.32)
3	$(p_{\text{tox}}, p_{\text{inf}})$	(0.05, 0.25)	(0.10, 0.10)	(0.30, 0.05)	(0.50, 0.02)	(0.60, 0.01)	—
	No. infused	2.6 (0.03)	1.2 (0.01)	0.5 (0.01)	0.0 (0.00)	0.0 (0.00)	—
	No. of toxicities	0.1 (0.00)	0.1 (0.00)	0.2 (0.01)	0.0 (0.00)	0.0 (0.00)	—
	% FMTD	0.0 (0.00)	0.4 (0.06)	2.8 (0.17)	0.4 (0.06)	0.0 (0.02)	**96.4 (0.19)**
4	$(p_{\text{tox}}, p_{\text{inf}})$	(0.01, 0.99)	(0.05, 0.95)	(0.07, 0.90)	(0.10, 0.75)	**(0.30, 0.50)**	—
	No. infused	1.0 (0.01)	1.5 (0.02)	7.3 (0.05)	8.5 (0.04)	**5.7 (0.04)**	—
	No. of toxicities	0.0 (0.00)	0.1 (0.00)	0.5 (0.01)	0.8 (0.01)	**1.7 (0.02)**	—
	% FMTD	0.0 (0.00)	0.4 (0.06)	9.2 (0.29)	28.8 (0.45)	**61.5 (0.49)**	0.1 (0.03)
5	$(p_{\text{tox}}, p_{\text{inf}})$	(0.50, 0.90)	(0.60, 0.75)	(0.70, 0.50)	(0.75, 0.25)	(0.80, 0.05)	—
	No. infused	3.2 (0.03)	3.6 (0.04)	1.7 (0.02)	0.0 (0.00)	0.0 (0.00)	—
	No. of toxicities	1.6 (0.01)	2.2 (0.02)	1.2 (0.01)	0.0 (0.00)	0.0 (0.00)	—
	% FMTD	4.8 (0.21)	3.2 (0.17)	0.0 (0.02)	0.0 (0.00)	0.0 (0.00)	**92.0 (0.27)**

since the patient is infused at $j = Y$ if $Y < j_{\text{CRM}}$ and at $j = j_{\text{CRM}}$ if $Y \geq j_{\text{CRM}}$, it follows that

$$\Pr(T = 1 \mid Y) = \Pr(T = 1 \mid j = Y)\Pr(Y < j_{\text{CRM}}) + \Pr(T = 1 \mid j = j_{\text{CRM}})\Pr(Y \geq j_{\text{CRM}}).$$

Under the first model given in Section 2.1 for the distribution of Y, this equals

$$p_Y^{\exp(\alpha)}\left(1 - q_{j_{\text{CRM}}}^{\exp(\beta)}\right) + p_{j_{\text{CRM}}}^{\exp(\alpha)} q_{j_{\text{CRM}}}^{\exp(\beta)}.$$

A more complex model might arise from the assumption that $\Pr(T = 1)$ depends not only on the dose level at which the patient is infused but, for biological reasons, on Y itself. In this case, e.g., $\Pr(T = 1 \mid j = 1, Y = 1) \neq \Pr(T = 1 \mid j = 1, Y = 2)$. An area for future investigation would be to explore this possibility based on data from a completed trial and, ideally, modeling biological explanations of such a relationship.

5. Simulations

In this section, we summarize results from a simulation study conducted while designing the DC-AL infusion trial. Given the five dose levels, the design parameters for this trial are $p^* = 0.30$ for toxicity, $\theta^* = 0.50$ for infusibility, $p_{\text{U,F}} = p_{\text{U,T}} = 0.90$, $N = 24$, $M = 48$, and $c = 2$. The simulations are summarized in Tables 1 and 2. Each scenario in the tables is characterized by fixed probabilities p_{tox} of toxicity and p_{inf} infusibility at each dose level, subject to the monotonicity constraints described earlier. These five scenarios were chosen from a larger set studied in order to provide a reasonable illustration of the design's properties. Numerical integrations required to compute mean posterior toxicity probabilities and the decision criteria 3 and 5 were done using a globally adaptive integrator choosing among several Gauss–Kronrod quadrature formulas, obtained from the Fortran subroutine QAG in the numerical integration package QUADPACK (Piessens et al., 1983). All computations were done on a DEC AlphaServer 4100 5/400 running Digital UNIX 4.0D. Each case was simulated 10,000 times, with a run time of about 25 minutes per case. The simulation program is written in Fortran 77 and Fortran 90 and utilizes a uniform random number generator based on the algorithm of L'Ecuyer and Cote (1991). The simulation program and a separate program for trial conduct are available from the second author on request.

Under scenario 1, $p_{\text{tox}}(d_2) = 0.30$, the CRM target, while $p_{\text{inf}}(d_j)$ is well above the specified lower limit of 0.50 for $j = 1, 2, 3$, and 4. Scenario 2 differs from scenario 1 only in that the infusibility probabilities are smaller, with $p_{\text{inf}}(d_3) = 0.50$ and higher doses not feasible. This has the effect of increasing the number of patients infused at d_2 and the likelihood of correctly selecting d_2 as the FMTD from 60.7% to 74.5% but also has the effect of slightly increasing the early-stopping (false negative) rate. Under scenario 3, while $p_{\text{tox}}(d_3) = 0.30$, the target, all $p_{\text{inf}}(d_j)$ values are well below 0.50. In this case, the algorithm correctly terminates the trial early 96.4% of the time and does so very quickly, after on average only 4.3 patients. Scenario 4 is a difficult case in that the highest dose level d_5 is best, but it is barely acceptable in terms of infusibility. The algorithm correctly selects d_5 as the FMTD 61.5% of

Table 2

Operating characteristics of the proposed CRM extension including the infusibility criterion (FCRM), the CRM ignoring infusibility and not infusing the patient if $Y < j_{\text{CRM}}$ (CRMa), and the CRM ignoring infusibility and infusing the patient at Y if $1 \leq Y < j_{\text{CRM}}$ (CRMb), where Y = dose level of cells grown. All three algorithms stop early if the lowest dose is unacceptably toxic.

		Scenario				
		1	2	3	4	5
No. enrolled	FCRM	22.8 (4.87)	24.4 (6.75)	18.1 (11.21)	24.2 (0.79)	10.4 (7.38)
	CRMa	25.1 (4.99)	38.3 (10.01)	48.0 (0.32)	39.1 (6.67)	20.3 (12.68)
	CRMb	22.8 (4.89)	24.3 (6.80)	20.5 (11.57)	24.2 (0.73)	10.4 (7.37)
No. infused	FCRM	22.5 (5.07)	21.8 (6.17)	4.3 (3.98)	24.0 (0.64)	8.6 (6.74)
	CRMa	22.9 (4.17)	21.4 (5.14)	2.3 (1.49)	23.7 (1.01)	12.7 (7.08)
	CRMb	22.6 (4.82)	21.9 (5.95)	5.1 (3.98)	24.0 (0.52)	9.4 (6.56)
% Correct decision	FCRM	60.7 (0.49)	74.5 (0.44)	96.4 (0.19)	61.5 (0.49)	92.0 (0.27)
	CRMa	52.9 (0.50)	51.6 (0.50)	0.0 (0.00)	74.1 (0.44)	79.3 (0.41)
	CRMb	59.8 (0.49)	73.7 (0.44)	0.0 (0.00)	61.2 (0.49)	92.2 (0.27)

the time. Most of the patients are treated at the upper three dose levels, where $0.07 \leq p_{\text{tox}} \leq 0.30$, and the algorithm selects d_4, for which $p_{\text{tox}} = 0.10$, in 28.8% of the runs. Under scenario 5, all of the doses are excessively toxic and, despite the fact that the first three levels are feasible, the algorithm correctly stops the trial early with 92% probability after treating on average only 8.6 patients. The algorithm could be made even safer by decreasing $p_{\text{U,T}}$, but this also has the effect of decreasing the correct selection rates under scenarios with acceptable dose levels.

We also simulated the design under model 2 with $a_+ = 2$, under the same scenarios using the same seed for random number generation. This yielded OCs virtually identical to those of model 1. It thus appears that, if the priors of the two models are calibrated so that each is uninformative and they have the same means, then they yield designs with nearly the same OCs. If a larger value of a_+ is used under model 2, however, this may change the OCs substantially under scenarios where feasibility is low. For example, if a_+ is increased from two to six, then the OCs remain essentially the same under scenarios 1, 2, 4, and 5 but change under scenario 3, where no dose level is feasible. In this case, the selection percentages for the five doses change from (0.0, 0.4, 2.8, 0.4, 0.0)% to (0.0, 4.3, 7.4, 2.0, 0.3)%, the average percent of trials stopped early drops from 96.4 to 86.0%, and the mean number of patients infused increases from 4.3 to 7.4. It thus seems advisable to use $a_+ = 2$ in practice with model 2, especially since our motivation is the possibility that the *ex vivo* laboratory methodology may not be feasible.

The following examples illustrate how the additional safety criterion protecting against excessive toxicity works in practice. Recall that the upper probability cutoff 0.90 is used for the early-stopping criterion $\Pr(p_j^{\exp(\alpha)} > 0.30 \mid \text{data})$ in the DC-AL trial. This criterion probability equals 0.86 if three toxicities are observed in the first four patients ($X_T(d_1)/n = 3/4$), so the trial would not be stopped in this case. The trial would be stopped if $X_T(d_1)/n = 4/4$, 4/5, or 4/6 since the criterion equals 0.93, 0.92 and 0.93, respectively, in these cases. Although patients are treated in cohorts of size 2, if, e.g., 4/5 toxicities are observed before the sixth patient is enrolled in the trial, then the trial is terminated without treating a sixth patient. More complex examples might involve patients being treated at dose levels above d_1 before the trial is stopped. While it is not uncommon in such circumstances for the investigators to add a dose level below d_1 rather than stop the trial, here and in similar dose-finding trials, we include the lowest dose levels that the investigators will consider. This ensures that the design and thus the OCs are a realistic reflection of actual trial conduct. If either the OCs or the decisions under cases such as those described above are not acceptable, then the design parameters may be calibrated appropriately. In the DC-AL infusion trial, no level below 10^6 cells per kg body weight was considered to be therapeutically acceptable, and it was planned from the start that, if the trial were terminated early due to excessive toxicity, all aspects of the therapeutic approach would be reevaluated and appropriately modified before beginning a subsequent trial.

To our knowledge, our proposal is the only method that formally addresses the problem that some patients may not be infusible at a given targeted dose in trials involving *ex vivo* expansion. Unfortunately, most dose-finding trials in oncology are conducted using some variant of the 3+3 algorithm, which, even when feasibility is not an issue, is well known to be inferior to the CRM. Consequently, although the 3+3 algorithm is actually current practice, it is not worthwhile to compare it with our proposed method. Two versions of the CRM that might be compared with our proposal would be to either infuse only patients for whom $Y \geq j_{\text{CRM}}$ (CRMa) or infuse patients for whom $1 \leq Y < j_{\text{CRM}}$ at Y and include their toxicity data in the CRM computations (CRMb). Table 2 summarizes properties of these two versions of the CRM and of our proposal, which we denote by FCRM. Although the safety criterion 5 is not typically used in the CRM, we included it in all three algorithms so that they differ only in how they deal with infusibility. Table 2 shows that FCRM and CRMb have roughly the same performance in all scenarios except 3, where, as one might expect, neither CRMa nor CRMb ever make the right decision. The only difference among the three methods under scenario 1 is the lower probability of a correct decision with CRMb. This is due in part to the fact

that CRMb is more likely to treat patients at dose levels with $p_{\text{tox}} > 0.30$ as the p_{inf} values decrease. Scenario 2 shows the cost of not infusing patients with $Y < j_{\text{CRM}}$ since CRMb requires 14 more patients enrolled to infuse the same number as do FMTD and CRMa. CRMb actually has the highest correct decision rate under scenario 4, although, again, at a cost of an additional 15 patients enrolled but not infused, and its performance is the worst in all regards under scenario 5.

6. Discussion

In recent years, various adoptive T-cell immunotherapy strategies have been studied in the treatment of brain tumors, kidney cancer, melanoma, ovarian cancer, and breast cancer, among other diseases. Although this is a relatively new type of treatment, clinical trials to date have established that the likelihood of both anticancer effects and adverse effects increase with the dose of cells infused. While there have been major advances with respect to the understanding of the biologic mechanisms underlying *ex vivo* T-cell expansion, a universally effective method remains to be established (Cheever et al., 1981). The problems common to all early-phase T-cell infusion trials are that a safe and effective dose must be determined in the presence of dose-limiting toxicity and the possibility that the number of cells generated *ex vivo* for a given patient may be less than a given desired level. Thus, we anticipate that our proposed methodology will be of use in numerous future trials.

More formal sequential methods than that proposed here are possible. A decision-theoretic Bayesian approach would require specification of a utility function, with its expectation maximized for each dose selection via backward induction. The computational limitations of such an approach are quite severe, however (Carlin, Kadane, and Gelfand, 1998). Alternatively, a frequentist test-based approach would require control of the overall probabilities of selecting correct and incorrect doses. Such formal methods necessarily require specification of numerical parameters, either to determine a Bayesian utility function or to determine frequentist test error probability limits. Thus, either of these methodologies, while nominally more objective than our approach, would be optimal only given such parameters. Exploration of such formal methods might be an area for future research.

We have defined the FMTD as an empirical quantity in rule 4 of the algorithm; i.e., the FMTD depends on posterior quantities and hence on the observed data. From a frequentist viewpoint, one could define the "true" MTD, based on a fixed dose–toxicity probability curve $F_T(d) = \Pr(T = 1 \mid d)$, as the unique dose $F_T^{-1}(p^*)$. Given a fixed probability $F_I(d)$ of infusibility at dose d, the true FMTD could then be defined as the dose closest to $F_T^{-1}(p^*)$ satisfying the inequality $F_I(d) \geq \theta^*$. Theoretical properties of the empirically determined FMTD might be explored under this frequentist definition of an FMTD.

An extension of the proposed algorithm would be required in dose-finding trials where both efficacy and toxicity, rather than toxicity alone, are used to determine an appropriate therapeutic dose. A statistical strategy for dose finding in such a setting has been proposed by Thall and Russell (1998). Under their formulation, patient outcome is trinary, taking on the possible values {response, toxicity, neither}, where "response" is a given efficacy event such as complete remission of leukemia or ≥50% shrinkage of a solid tumor, with the events defined so that response and toxicity are disjoint. In this case, their dose-finding algorithm could be modified in a reasonably straightforward manner to accommodate feasibility similarly to the algorithms proposed here.

Acknowledgement

This research was partially supported by National Cancer Institute grant R01 CA83932.

Résumé

Une nouvelle forme de traitement du cancer consiste en l'administration au patient de cellules-T spécifiquement dirigées contre la tumeur. Ces cellules-T sont prélevées sur le patient et multipliées ex-vivo. Le but thérapeutique est la destruction sélective des cellules cancéreuses par ces cellules-T "activées". Un problème important dans les phases précoces du développement d'un anticancéreux, avant d'entreprendre un essai clinique de phase II, est la détermination de la dose maximale tolérée (MTD = maximal tolerated dose). La dose peut être évaluée par le nombre de cellules administrées par unité de poids corporel et la MTD définie à partir de la probabilité de manifestations toxiques en fonction de la dose. Comme pour un essai de phase I d'une nouvelle chimiothérapie anticancéreuse, cela peut être fait en administrant différentes doses à des groupes successifs de patients, chaque nouvelle dose étant choisie en fonction des manifestations toxiques constatées sur les patients précédemment traités. Une telle méthode de recherche de dose est inadaptée aux essais de traitement par cellules-T car le nombre de cellules cultivées ex-vivo, pour un patient donné, peut ne pas être disponible en quantité suffisante pour la dose qu'on souhaite lui administrer. Pour traiter ce problème nous présentons un algorithme pour la conduite de l'essai qui détermine une MTD accessible basée à la fois sur les probabilités de toxicité et de disponibilité. La méthode est illustrée par une application à un essai de traitement de leucémie aiguë par administration de lymphocytes activés. Une étude de simulation montre que la méthodologie proposée est à la fois sûre et fiable.

References

Ahn, C. (1998). An evaluation of phase I cancer clinical trial designs. *Statistics in Medicine* **17,** 1537–1549.

Alyea, E. P., Soiffer, R. J., Canning, C., Neuberg, D., Schlossman, R., Pickett, C., Collins, H., Wang, Y., Anderson, K. C., and Ritz, J. (1998). Toxicity and efficacy of defined doses of $CD4^+$ donor lymphocytes for treatment of relapse after allogeneic bone marrow transplant. *Blood* **91,** 3671–3690.

Babb, J., Rogatko, A., and Zacks, S. (1998). Cancer phase I clinical trials: Efficient dose escalation with overdose control. *Statistics in Medicine* **17,** 1103–1120.

Carlin, B. P., Kadane, J. B., and Gelfand, A. E. (1998). Approaches for optimal sequential decision analysis in clinical trials. *Biometrics* **54,** 964–975.

Cheever, M. A., Greenberg, P., and Fefer, A. (1981). Specific adoptive immunotherapy of established leukemia with synergistic lymphocytes sequentially immunized *in vivo* and *in vitro* and non-specifically expanded by culture with interleuken-2. *Journal of Immunology* **126,** 1318–1322.

Choudhury, A., Gajewski, J. L., Liang, J. C., Popat, U., Claxton, D. F., Kliche, K. O., Andreeff, M., and Champlin, R. E. (1997). Use of leukemic dendritic cells for the generation of antileukemic cytotoxicity against Philadelphia chromosome positive chronic myelogenous leukemia. *Blood* **89,** 1133.

Choudhury, A., Liang, J. C., Thomas, E. K., Flores-Ramo, L., Xie, Q. S., Agusala, K., Sutaria, S., Sinha, I., Champlin, R. E., and Claxton, D. F. (1999). Dendritic cells derived *in vitro* from acute myelogenous leukemia cells stimulate autologous, antileukemic T-cell responses. *Blood* **93,** 780.

Dougherty, T. B., Porsche, V., and Thall, P. F. (1999). Maximal tolerated dose of nalmefene in patients receiving epidural fentanyl and dilute bupivacaine for postoperative analgesia. *Anesthesiology* **92,** 1010–1016.

Figlin, R. A., Pierce, W. C., Kaboo, R., Tso, C. L., Moldawer, N., Gitlitz, B., deKernion, J., and Belldegrun, A. (1997). Treatment of metastatic renal cell carcinoma with nephrectomy, interleuken-2 and cytokine-primed or CD-8 selected tumor-infiltrating lymphocytes from primary tumor. *Journal of Urology* **158,** 740–745.

Freedman, R. S., Edwards, C. L., Kavanagh, J. J., et al. (1994). Intraperitoneal adoptive immunotherapy of ovarian carcinoma with tumor-infiltrating lymphocytes and low dose recombinant interleuken-2: A pilot trial. *Journal of Immunotherapy and Tumor Immunology* **16,** 198–210.

Goodman, S. N., Zahurak, M. L., and Piantadosi, S. (1995). Some practical improvements in the continual reassessment method for phase I studies. *Statistics in Medicine* **14,** 1149–1161.

L'Ecuyer, P. and Cote, S. (1991). Implementing a random number package with splitting facilities. *ACM Transactions on Mathematical Software* **17,** 98–111.

Møller, S. (1995). An extension of the continual reassessment methods using a preliminary up-and-down design in a dose finding study in cancer patients, in order to investigate a greater range of doses. *Statistics in Medicine* **14,** 911–922.

O'Quigley, J. and Shen, L. Z. (1996). Continual reassessment method: A likelihood approach. *Biometrics* **52,** 673–684.

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.

Piantadosi, S., Fisher, J. D., and Grossman, S. (1998). Practical implementation of a modified continual reassessment method for dose-finding trials. *Cancer Chemotherapy and Pharmacology* **41,** 429–436.

Piessens, R., deDoncker-Kapenga, E., Uberhuber, C., and Kahaner, D. (1983). *Quadpack: A Subroutine Package for Automatic Integration,* Volume 1, *Series in Computational Mathematics.* New York: Springer Verlag.

Plautz, G. E., Barnett, G. H., Miller, D. W., Cohen, B. H., Prayson, R. A., Krauss, J. C., Luciano, M., Kangisser, D. B., and Shu, S. (1998). Systematic T cell adoptive immunotherapy of malignant gliomas. *Journal of Neurosurgery* **89,** 42–51.

Porter, D. L., Roth, M. S., McGarigle, C., Ferrera, J. L., and Antin, J. H. (1994). Induction of graft-versus-host disease as immunotherapy for relapsed chronic myelogenous leukemia. *New England Journal of Medicine* **330,** 100–106.

Thall, P. F. and Russell, K. E. (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.

Thall, P. F., Lee, J. J., Tseng, C.-H., and Estey, E. H. (1999). Accrual strategies for phase I trials with delayed patient outcome. *Statistics in Medicine* **18,** 1155–1169.

Received October 2000. *Revised April* 2001.
Accepted April 2001.