Featured Article

Patient Characteristics Compete with Dose as Predictors of Acute Treatment Toxicity in Early Phase Clinical Trials

André Rogatko,[1] James S. Babb,[1] Hao Wang,[1] Michael J. Slifker,[1] and Gary R. Hudes[2]

Departments of [1]Biostatistics and [2]Medical Oncology, Fox Chase Cancer Center, Philadelphia, Pennsylvania

ABSTRACT

Purpose: **The purpose of this study was to identify patient characteristics that may be risk factors or markers of susceptibility to adverse treatment effects in cancer Phase I and II clinical trials.**

Patients and Methods: **A total of 459 patients enrolled in 23 therapeutic Phase I and II studies at the Fox Chase Cancer Center were included in the analysis. Patient-specific characteristics, medical and treatment history, doses of experimental agents, and graded toxicities were extracted from case report forms. We developed a novel summary measure, the toxicity index (TI), to better discriminate patients on the basis of their overall toxicity experiences. Mixed model ANOVA was used to model TI on the basis of data from all trials using a specific agent. Generalized estimating equations in the context of binary logistic regression were used to model dose-limiting toxicity.**

Results: **Seventeen pretreatment factors, including performance status, alkaline phosphatase, total bilirubin, serum creatinine, and tobacco use, emerged as significant predictors of toxicity as defined by dose-limiting toxicity or TI. Unexpectedly, dose was not always a predictor of toxicity. Even for values within the normal range, the TI identified serum bilirubin and alkaline phosphatase as predictors of toxicity after treatment with docetaxel and alkaline phosphatase as a predictor for toxicity after treatment with irinotecan.**

Conclusions: **Independent of dose, certain pretreatment characteristics, including measures of organ function that are in the normal range, were found to be predictors of treatment toxicity. Because of its sensitivity to differences in overall toxicity, the TI should prove to be a useful tool for identifying predictors of chemotherapy-related toxicity.**

INTRODUCTION

Clinical trials of new anticancer therapies are essential tools in the search for more effective cancer treatment. Cancer clinical trials typically proceed through distinct phases. The major objective in Phase I trials is to identify an optimal dose for subsequent Phase II and III studies, in which the main goal is an evaluation of treatment efficacy. The dose for Phase II trials is typically defined in terms of iatrogenic adverse events observed in Phase I and represents the dose expected to induce unacceptable toxicity in a specified fraction of patients. Generally, the Phase II dose is determined on the basis of data from a relatively limited number of patients. Other than the extent of prior treatment, the Phase II dose does not take into account identifiable patient differences that could affect treatment tolerance. That is, the Phase II dose is not adjusted on the basis of measurable patient characteristics that may be predictive of adverse responses to treatment.

Several studies have examined data retrospectively from early phase clinical trials to assess treatment response rates or evaluate the adequacy of conventional methods for designing clinical trials. For example, Decoster *et al.* (1) summarized the antitumor activity and toxic deaths reported in single-agent Phase I trials using cytotoxic compounds published between 1972 and 1987, whereas Dillman and Koziol (2) reviewed the objective response rates in Phase I trials involving 6447 patients. Smith *et al.* (3) analyzed the data from all published Phase I trials conducted at the M. D. Anderson Cancer Center from 1991 to 1993 to assess the adequacy of using the standard modified Fibonacci method to determine the dose recommended for Phase II study. Each of these studies concluded that there was a need for alternative and improved methods of designing, conducting, and interpreting Phase I clinical trials. This overall conclusion was supported by Eisenhauer *et al.* (4).

Various studies have examined prognostic factors for survival as an end point in early phase cancer clinical trials. For example, Yamamoto *et al.* (5) examined the survival of 82 lung cancer patients treated in Phase I trials at the Japan National Cancer Center Hospital between 1987 and 1993. Their multivariable survival analysis identified performance status > 1, weight loss $> 10\%$, and number of metastatic sites > 1 as risk factors. Similarly, using data from 154 Phase I patients treated at Centre Leon Berard (Lyon, France), Bachelot *et al.* (6) identified performance status 2 or 3 ($P < 0.001$) and lactate dehydrogenase levels greater than 600 IU ($P < 0.001$) as independent predictors of overall survival. These studies demonstrate that subgroups with different survival expectancy may be identified among patients who participate in cancer Phase I clinical trials.

Unlike survival and antitumor activity, we found no systematic study of prognostic factors for the end point of toxicity.

Received 11/5/03; revised 2/4/04; accepted 2/18/04.

Grant support: Grants CA06927 and CA-92769 from the National Cancer Institute, a Tobacco Formula Grant from the State of Pennsylvania, and an appropriation from the Commonwealth of Pennsylvania. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: A. Rogatko is currently at Winship Cancer Institute, Emory University School of Medicine, 1365 C Clifton Road, NE, Atlanta, GA 30322.

Requests for reprints: André Rogatko, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111. Phone: (215) 728-2450; Fax: (215) 728-2553; E-mail: a_rogatko@fccc.edu.

The aim of the present study was to identify readily available patient characteristics that may constitute risk factors or markers of susceptibility to adverse treatment effects in cancer Phase I and II clinical trials. Because the factors considered in this study can be observed before treatment, they may form the foundation for customized dosing regimens wherein drug doses are adjusted according to individual patient susceptibilities.

PATIENTS AND METHODS

Patient Population. A total of 459 patients enrolled in 23 Fox Chase Cancer Center therapeutic Phase I and II studies during the time period 1991–1999 were included in the analysis. The Fox Chase Cancer Center Research Review Committee and Institutional Review Board approved each study. Before screening and protocol enrollment, each patient provided written informed consent. In general, normal or near normal renal and hepatic function were required for these early clinical trials, although the specific criteria varied.

Data Extraction. Patient-specific characteristics, medical and treatment history, doses of the experimental agents, and toxicities were extracted from case report forms and entered into an Excel spreadsheet. For each trial, the current version of the National Cancer Institute Common Toxicity Criteria (CTC) were utilized to grade all toxicities, and the type and grade of worst toxicity experienced by each patient were recorded. The toxicities for this analysis were considered definitely, probably, or possibly related to treatment.

Covariates. The following patient-specific factors were considered as candidate predictors of adverse treatment response, each determined at or before treatment onset: gender; age at treatment onset; age at diagnosis; stage at diagnosis; Eastern Cooperative Oncology Group performance status; weight (kg) at treatment onset; body surface area; serum total bilirubin (mg/dl); albumin (g/dl); alkaline phosphatase (units/liter); aspartate aminotransferase (units/liter); alanine aminotransferase (units/liter); serum creatinine (mg/dl); disease site and type; prior chemotherapy; cumulative radiation to bone marrow; and binary indicators for recent weight loss (*i.e.*, $\geq$10% loss) or anorexia; alcohol use; tobacco use; or history of cardiovascular disease. In addition to these patient-specific factors, the following were included in the model as covariates: the phase (Phase I, II, I/II) of investigation; and the dose of each experimental agent, expressed either in mg, mg/m^2, mg/kg, or area under the curve (AUC).

Toxicity. Toxicity data are presented for each subject as a set of treatment-attributable adverse events, each graded (0, 1, 2, 3, or 4) according to National Cancer Institute CTC. The term "toxicity profile" will be used to refer to the collection of all graded toxicities observed for an individual patient. To include all trials in the same analysis, we adopted a common definition of dose-limiting toxicity (DLT) that was consistent with the definitions of DLT used in the individual trials. Specifically, DLT was defined as occurring whenever any one of the following was manifested: grade 4 hematological toxicity (grade 4 neutropenia or thrombocytopenia of any duration); or a grade 3 or 4 nonhematological toxicity (excluding nausea, vomiting, and alopecia); and febrile neutropenia (grade 3 or 4).

To identify risk factors and other pretrial characteristics of interest, we considered several ways of summarizing the toxicity profile of a given patient into a single measure of overall severity. Simple measures might include a binary indicator of DLT and the maximum toxicity grade experienced by the patient. Although easy to interpret and compute, these measures lack the ability to discriminate between patients on the basis of their overall toxicity profiles. For example, a subject experiencing five distinct grade 3 nonhematological toxicities has the same toxicity score, with respect to either measure, as one with a single grade 3 nonhematological toxicity. Consequently, we developed a new summary measure that takes into account all observed toxicity grades rather than just the most severe one. The summary measure, which we refer to as the toxicity index (TI), is computed according to the following algorithm: let the toxicity grades in a subject's toxicity profile be represented in descending order by the sequence $X_1 \geq X_2 \geq \ldots \geq X_n$. Calculate the subject's TI score as the weighted sum of the ordered toxicity grades: $TI = \sum_{i=1}^{n} w_i X_i$, where the weights are given by

$$w_i = \prod_{j=1}^{i-1} (X_j + 1)^{-1}. \quad \text{(A)}$$

Specifically, the TI is calculated as

$$TI = X_1 + \frac{X_2}{1 + X_1} + \frac{X_3}{(1 + X_1)(1 + X_2)} + \cdots + \frac{X_n}{(1 + X_1)\cdots(1 + X_{n-1})}. \quad \text{(B)}$$

For example, a subject exhibiting 2 grade 3 toxicities will have a score of TI = 3 + 3/4 = 3.75, whereas a subject with one grade 3 and ten grade 2 toxicities will have a score of TI = $3 + 2/4 + 2/(3 \cdot 4) + 2/(3^2 \cdot 4) + \cdots + 2/(3^9 \cdot 4) \approx 3.74999$.

The TI can be generalized to accommodate the differential impact of various toxicities by applying relative weights or appropriate transformations to the CTC graded toxicities before the application of the algorithm to compute the TI. For example, we considered grade 3 neutropenia, leukopenia, and anemia non-dose-limiting, and we re-coded these as the observed CTC grade minus 1 (to a minimum of 0 if no toxicity was present). Thus, uncomplicated grade 3 neutropenia was downgraded to a score of 2. This *ad hoc* adjustment is consistent with the convention in Phase I trials of not considering uncomplicated grade 3 neutropenia as a DLT. It allows all grade 3 nonhematological toxicity, but not grade 3 (actual grade) hematological toxicity other than thrombocytopenia, to be considered a DLT. In this way, the TI conveys the information provided by the binary indicator of DLT while also permitting two patients who both do or do not express DLT to be distinguishable if they have different overall toxicity profiles. By design, the TI has several properties of interest.

1. Any score greater than or equal to 3 corresponds to the DLT definition given above, and the maximum toxicity grade is the integer part of the final score. For example, a TI = 3.0 indicates a single grade 3 toxicity, whereas a score of 3.5 indicates that the patient experienced at least one grade 3

toxicity plus additional toxicity. Hence, the TI preserves the highest toxicity grade used to define DLT.
2. All toxicity grades are taken into account, although lower grades will tend to contribute less to the final score.
3. The score is a number between 0 and 5.
4. A large number of toxicities of the same grade will generate a TI score just slightly less than that generated by a single toxicity of the next higher grade.

Statistical Analysis. Mixed model ANOVA was used to model the TI using data from all trials involving a specific agent. The model to predict the TI included the dose of agent received by each patient as a covariate and trial institutional review board number as a random classification variable (to adjust for systematic differences between different clinical trials using the same agent). Generalized estimating equations in the context of binary logistic regression were used to model DLT. Because patients in any single trial were similar in terms of treatment-related factors such as infusion time, mode of administration, and general dose range, a variable (protocol identifier) was created to distinguish this group or cluster of patients from those participating in other trials. The inclusion of this variable as a classification factor in the analysis permitted an adjustment for factors that changed from trial to trial but exhibited negligible variation within a single trial. Similarly, a variable was created to distinguish between Phase I and Phase II patients. Together with dose, the indicators of study phase (Phase I *versus* II) and protocol (institutional review board number) were included in the model as covariates, with dose and phase as fixed effects and protocol as a random effect. Backward variable selection was performed to identify a subset of covariates that best predicts toxic response (*i.e.*, TI and DLT) to each agent. A covariate was selected as a significant predictor if its comparison-wise *P* from the mixed model analysis was less than 0.05. Type III sum of squares were used for hypothesis testing. The adequacy of the fitted model was assessed with diagnostic procedures. Residual plots and normal probability plots of residuals were used.

RESULTS

The characteristics of the 459 patients included in the analysis are summarized in Tables 1 and 2. The median age was 62 years (range, 27–84 years). The analysis of toxicity was restricted to agents for which there was adequate toxicity data, *i.e.*, more than 30 patients treated with each drug. As reflects Phase I and early Phase II clinical trial populations, the baseline renal and hepatic functions were either normal or minimally abnormal, *i.e.*, considered as grade 0–1 elevations by the NIH CTC versions 1.0 and 2.0, and performance status was usually 0–1 on the Eastern Cooperative Oncology Group scale. The agents used most frequently were paclitaxel, carboplatin, and estramustine, consistent with the diseases treated most often on Phase I and II trials at our institution (non-small cell lung, prostate, head and neck, and ovarian cancer).

We first determined which dose expression for each agent (mg, mg/m^2, mg/kg, or AUC) was the best predictor of the two toxicity end points, TI (as a continuous variable) and DLT (as a binary variable). The two-sided *P* values of each expression are reported in Table 3. These values indicate the strength of association of either DLT or TI with drug dose as a single variable.

Table 1 Absolute and relative frequencies

A. Absolute and relative frequencies of 459 patients by phase of clinical trial, gender, and agent

Variable	Category	Patients	Percentage
Phase	I	275	59.9
	I/II	24	5.2
	II	160	34.9
Gender	F	153	33.3
	M	306	66.7
Agent	Paclitaxel	245	53.4
	Estramustine	154	33.6
	Carboplatin	140	30.5
	Cisplatin	45	9.8
	Docetaxel	48	10.5
	Irinotecan	36	7.8
	Tomudex	36	7.8
	Tipifarnib (R115777)[a]	34	7.4
	5-Flourouracil	30	6.5
	Gemcitabine	19	4.1
	Topotecan	14	3.0
	Cytoxan	13	2.8
	Vinblastine	12	2.6
	Bryostatin	12	2.6
	BMS-188797[b]	11	2.4
	BMS-214662[a]	11	2.4

B. Absolute and relative frequencies of 459 patients by highest toxicity grade

Highest toxicity			
Overall	No toxicity	10	2.2
	1	84	18.3
	2	116	25.3
	3	139	30.3
	4	110	24.0
Nonhematological	No toxicity	43	9.4
	1	147	32.0
	2	121	26.4
	3	110	24.0
	4	38	8.3
Hematological	No toxicity	114	24.8
	1	68	14.8
	2	105	22.9
	3	80	17.4
	4	92	20.0

C. Absolute and relative frequencies of 459 patients by primary disease site, prior chemotherapy or radiotherapy, performance status, alcohol use, cigarette smoking, and weight loss

Variable	Category	Patients	Percentage
Primary disease site	Prostate	117	25.55
	Other	103	22.49
	Lung	73	15.94
	Ovary	42	9.17
	Colon	41	8.95
	Esophageal	39	8.52
	NHL	20	4.37
	Pancreas	14	3.06
	Breast	9	1.97
Prior radiation	<25%	453	99.34
	>25%	3	0.66
Prior chemotherapy	No	226	49.24
	Yes	233	50.76
Performance status	0	177	39.16
	1	252	55.75
	2	23	5.09
Alcohol use	No	182	45.61
	Yes	217	54.39
Cigarette smoking	No	168	41.38
	Yes	238	58.62
Weight loss	No	344	76.27
	Yes	107	23.73

[a] Farnesyltransferase inhibitor.
[b] Taxol analog.

Table 2 Summary statistics for baseline alkaline phosphatase, serum total bilirubin, albumin, SGOT,[a] and serum creatinine

Covariate	Mean ± SD	Median	25% quantile	75% quantile
Alkaline phosphatase	158.9 ± 185.6	105.0	80	157
Serum total bilirubin	0.58 ± 0.23	0.6	0.4	0.7
Albumin	3.78 ± 0.45	3.8	3.5	4.1
SGOT	28.1 ± 16.6	24.0	19.0	32.0
Serum creatinine	0.93 ± 0.26	0.9	0.8	1.1

[a] SGOT, aspartate aminotransferase.

Although there were some differences in correlation for each of the expressions of dose with the two toxicity end points (*e.g.*, for tipifamib), these differences were usually small. This univariate analysis showed that carboplatin and docetaxel doses were not strongly associated with either DLT or TI, whereas associations of drug dose with either DLT, TI, or both were found for each of the other agents. Expressions shown in bold in Table 3 were chosen as the units of drug dose for the subsequent analyses of covariates.

Table 4 shows the association of toxicity with those covariates observed to be significant predictors of toxic response to at least one agent. For each covariate, *P* values from the mixed model analyses and from generalized estimating equation analyses are presented. These *P* values assess the effect of each covariate when considered individually and adjusted for drug dose. Covariate *P* values significant at a comparisonwise 5% level have been set in bold in Table 4, as illustrated for docetaxel, where pretreatment alkaline phosphatase and total bilirubin were both significantly correlated with TI, and alkaline phosphatase was also correlated with occurrence of DLT. For all of the drugs studied, the DLT and TI analyses yielded a total of 17 potential covariates of toxicity. DLT alone as a measure of toxic response identified nine potential covariates, of which three were not identified by the TI. By contrast, the TI identified 14 potential covariates, including 7 not identified by considering only DLT. Based on these results, all the individually significant covariates for either TI or DLT were included in each agent's model, adjusted for phase of trial and dose expression. This final analysis, with the results summarized in Table 5, selected the "best subset" of covariates for each agent. *P* values of the predictors significant at a comparisonwise 5% level are shown in bold in Table 5. It is worth noting that for every agent considered, at least one pretreatment patient-specific characteristic was found to be a significant predictor of treatment toxic response. The adequacy of the fitted model was assessed with diagnostic procedures (residual plots and normal probability plots of residuals), and the diagnostic tools indicated that the underlying model assumptions were valid.

We found that dose did not correlate significantly with TI or DLT for carboplatin, tomudex, or docetaxel, whereas patient pretreatment total bilirubin and alkaline phosphatase correlated with DLT, TI, or both for all three agents. Alkaline phosphatase also correlated with TI for cisplatin, estramustine, and tipifarnib (R115777). The TI identified performance status as a predictor of toxicity, but not DLT, for paclitaxel and estramustine. Other factors predictive of DLT or TI were weight loss (for paclitaxel), creatinine (for the combination of tomudex and irinotecan), and tobacco use, which correlated with TI and DLT after treatment with irinotecan and tomudex. Interestingly, the correlation of total bilirubin with TI for docetaxel occurred over the normal range of serum bilirubin values, not just values above the upper limit of normal (Fig. 1).

Figs. 1 and 2 show, for selected agents, scatter plots of TI in the *ordinates* by selected significant predictors in the *abscissas*. For each set of pairs of observations from a given trial, an *ellipse* was drawn to enclose 95% of the distribution (95% quantile) under the assumption of a bivariate normal distribution. The shape of the ellipse depends on the correlation coefficient between TI and the specific patient characteristic. In each case, the major axis of the ellipse has a positive slope, indicating that higher levels of the relevant covariate were associated with more severe toxicity, and the ellipse was relatively broad, indicating that the association was only moderate in strength (approximately 25% of the variance in TI was explained by each covariate).

DISCUSSION

The basis for the variable toxicity experienced by patients in clinical trials of chemotherapeutic agents is largely unknown. Conventional approaches to Phase I trial design assume minimal interindividual pharmacokinetic and pharmacodynamic variability and assume that drug dose will be a significant determinant of toxicity. A uniform study population for Phase I trials is sought by selecting patients with normal organ function, good performance status, and minimal prior treatment. However, there are many other potential determinants of toxicity, just as there are unknown determinants of antitumor effect for most chemotherapeutic agents. Both genetic (*e.g.*, CYP450 profile) and environmental factors (*e.g.*, hepatic and renal function) are likely to appear in a more complex, but realistic, predictive model of host toxic response to chemotherapy.

The goal of this systematic toxicity analysis was to gener-

Table 3 Dose expressions (mg, mg/m^2, mg/kg, and AUC[a]) as predictors of toxicity end points: TI and DLT for selected agents

P values of each expression are reported. Dose expressions in bold were chosen as the expression of the dose for the subsequent analyses on the effect of covariates.

Agent	Outcome	mg	mg/m^2	mg/kg	AUC
Paclitaxel	TI	**0.0151**	0.0370	0.2124	
	DLT	0.8237	0.6095	**0.5493**	
Estramustine	TI	0.0627	**0.0484**	0.0626	
	DLT	0.3102	0.255	**0.2197**	
Carboplatin	TI	0.3482	0.6281	0.8771	**0.3146**
	DLT	0.4303	0.3419	0.3089	**0.2937**
Cisplatin	TI	<0.0001	**<0.0001**	<0.0001	
	DLT	0.2874	**0.2889**	0.3107	
Docetaxel	TI	0.7426	0.3894	**0.2863**	
	DLT	0.9485	0.2870	**0.0644**	
Irinotecan	TI	0.8385	0.1276	**0.0204**	
	DLT	0.9833	0.3322	**0.1831**	
Tomudex	TI	0.5114	0.1508	**0.0438**	
	DLT	0.2440	0.1489	**0.0961**	
Tipifarnib	TI	0.1891	**0.0368**	0.1730	
	DLT	0.2005	**0.0328**	0.0582	

[a] AUC, area under the curve; TI, toxicity index; DLT, dose-limiting toxicity.

Table 4 Individual covariates as predictors of toxicity end points, TI,[a] and DLT

Covariate *P* values significant at a comparisonwise 5% level are in bold.

Covariate	Outcome	Paclitaxel (mg)	Estramustine (mg/m^2)	Carboplatin (AUC)	Cisplatin (mg/m^2)	Docetaxel (mg/kg)	Irinotecan (mg/kg) + tomudex (mg/kg)	Tipifarnib (mg/m^2)
Alkaline phosphatase	TI	0.0790	**0.0072**	0.2597	**0.0334**	**0.0324**	0.0748	**0.0187**
	DLT	0.3959	0.0725	0.4501	0.0981	**0.0308**	**0.0209**	**0.0430**
Serum total bilirubin	TI	0.4284	0.3750	**0.0050**	0.2821	**0.0001**	0.4389	0.6609
	DLT	0.2037	0.1065	0.0799	0.1909	0.1499	0.2804	0.3888
Albumin	TI	0.0678	**0.0010**	0.9498	0.9491	0.7258	0.2180	**0.0338**
	DLT	0.0700	0.0635	0.2590	0.5556	0.6582	0.0525	**0.0117**
SGOT	TI	0.2356	**0.0039**	0.6463	0.5991	0.1813	0.7484	0.1226
	DLT	0.8915	**0.0333**	0.9852	0.2121	0.0724	0.3047	0.1525
Serum creatinine	TI	0.7484	0.5231	0.6889	0.7863	0.9615	**0.0290**	0.6573
	DLT	0.4070	0.4024	0.8538	0.6291	0.9471	0.1208	0.5766
Performance status	TI	**<0.0001**	**<0.0001**	0.7604	0.2047	0.2561	0.2399	0.1539
	DLT	0.0811	**0.0561**	0.8371	0.2571	0.1992	0.2387	0.1071
Weight loss	TI	0.0623	0.1133	0.8882	0.1249	0.0816	**0.0434**	0.8505
	DLT	**0.0398**	0.0858	0.3470	0.1146	0.4490	**0.0231**	0.7976
Cigarette smoking	TI	0.6733	0.9830	0.9531	0.2219	0.4066	**0.0111**	0.9465
	DLT	0.5616	0.1667	0.3533	0.4362	0.7965	0.1590	0.5606
Prior chemotherapy	TI	0.6330	0.2347	0.9345	0.5697	0.8315	0.2155	0.5675
	DLT	0.055	0.058	0.4158	0.4772	0.5313	0.0314	0.6960

[a] TI, toxicity index; DLT, dose-limiting toxicity; AUC, area under the curve; SGOT, aspartate aminotransferase.

ate hypothesis for the investigation of factors predictive of chemotherapy toxic response. Our intention was not to validate each hypothesis but to generate hypotheses that may be tested prospectively in subsequent clinical trials. Clinical experience provides the basis to decide which, if any, findings might be worthy of further investigation. Because we provide an assessment of the impact of statistically significant covariates on the probability of toxic response (Table 5), clinical investigators should be able to focus their attention on those factors observed to influence treatment susceptibility to a clinically relevant degree. This will not reduce the number of false or clinically misleading positives, but it will at least reduce the number of covariates that must be considered in subsequent confirmation trials to those that the present data suggest are influential.

Caution must be exercised when comparing results obtained for TI and DLT because different underlying analyses

Table 5 Selected "best subset" of covariates for each agent

Estimated COEF,[a] SEs, and *P* values

Agent	Covariate	Toxicity index			DLT		
		COEF	SE	*P*	COEF	SE	*P*
Paclitaxel	Dose (mg)	−0.00196	0.00092	**0.0341**	0.0019	0.004	0.6458
	Performance status	0.4268	0.1074	**0.0001**			
	Weight loss				1.0274	0.189	**0.0182**
Estramustine	Dose (mg/m^2)	0.000539	0.00017	**0.0041**			
	Alkaline phosphatase	0.000827	0.000296	**0.0059**			
	Performance status	0.6213	0.1233	**0.0001**			
Carboplatin	Dose (AUC)	0.0436	0.0508	0.4060			
	Serum total bilirubin	1.1066	0.3941	**0.0057**			
Cisplatin	Dose (mg/m^2)	−0.0402	0.0068	**0.0001**			
	Alkaline phosphatase	0.0065	0.0024	**0.0111**			
Docetaxel	Dose (mg/kg)	0.6648	0.3945	0.0997	1.257	1.206	0.2850
	Alkaline phosphatase				0.0161	0.0075	**0.0032**
	Serum total bilirubin	2.7418	0.6544	**0.0001**			
Irinotecan + tomudex	Irinotecan dose (mg/kg)	1.535	0.3917	**0.0006**	3.418	2.093	**0.0439**
	Tomudex dose (mg/kg)	11.57	7.39	0.1301	11.89	26.04	0.6465
	Alkaline phosphatase	0.00623	0.00156	**0.0005**	0.0201	0.0081	**0.0009**
	Serum creatinine	1.83	0.61	**0.006**			
	Cigarette smoking	−1.41	0.33	**0.0003**	−3.42	1.52	**0.0067**
Tipifarnib	Dose (mg/m^2)	0.0052	0.0025	**0.0495**	0.0115	0.0092	0.1869
	Alkaline phosphatase	0.00205	0.00082	**0.0187**	0.0183	0.011	**0.0314**
	Albumin				−4.93	2.3	**0.0030**

[a] COEF, coefficients; DLT, dose-limiting toxicity; AUC, area under the curve.

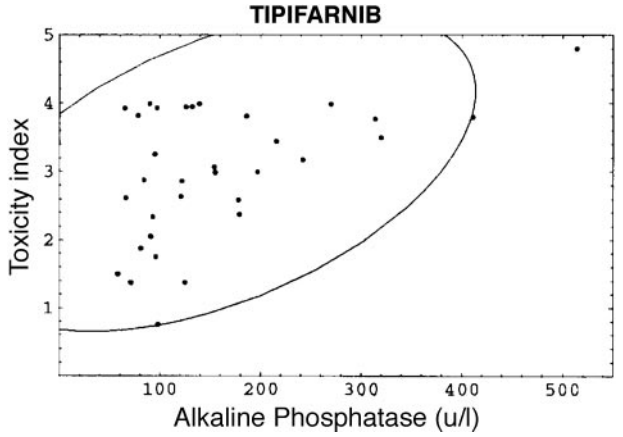

Fig. 1 Toxicity index (*Y* axis) by alkaline phosphatase (units/liter; *X* axis) in patients treated with tipifarnib (R115777). *Dots* are observed values for IRB97030, the *ellipse* encloses 95% of the distribution, and the Pearson correlation coefficient is $r = 0.501$ with $P = 0.003$.

were performed (mixed model ANOVA *versus* generalized estimating equations). However, the TI appeared to be more sensitive than the presence or absence of DLT (usually defined as a single grade 3–4 toxicity) alone to uncover pretreatment factors predictive of toxicity in the sense that more predictors were identified by the TI compared with DLT (14 *versus* 9). Although the role of each of these potential toxicity predictors requires validation in an independent sample, our results suggest that the TI may be a valuable tool to identify potential predictors of toxicity.

The TI differs from DLT by including all toxicities and all grades of toxicity. In most Phase I studies, a single grade 3 or 4 toxicity such as stomatitis or creatinine elevation would constitute DLT and would not be distinguished from two or more toxicities in the same patient for the purpose of dose-finding. By contrast, the TI includes both of these adverse events and assigns a higher score to the occurrence of two or more grade 3 toxicities as compared to only one grade 3 toxicity in a single patient. The TI also captures grade 1 and 2 toxicities, so that a patient with grade 2 stomatitis, grade 2 creatinine elevation, and grade 4 thrombocytopenia has a higher TI than one with only the latter two toxicities (TI of 3.67 *versus* 3.5).

Although grade 1 and 2 toxicity is usually not considered "dose-limiting" in the context of dose-finding for drug development, these adverse events reflect tissue damage and make long-term treatment difficult for many patients, particularly when multiple toxicities occur simultaneously. The TI weights grade 1 and 2 toxicity in proportion to the more severe grades, so that a patient with two grade 2 toxicities has a TI lower than that of a patient with a single grade 3 toxicity but a higher TI than that of a patient with only a single grade 2 toxicity (TIs of 2.67 *versus* 3.0 *versus* 2.0). Thus, the TI is consistent with the notion that both number and grade of toxicities are important and that multiple toxicities are likely to affect a patient to a greater extent than one toxicity. The TI describes a continuous range of toxic responses over which to establish correlation with patient characteristics. Like all summary measures, the major limitation of the TI is loss of relation to the individual components, that is, the individual toxicities and their grades.

We confirm previously reported correlations of hepatic function with toxicity of irinotecan and docetaxel (7–9). For both agents, dose modification has been recommended for total bilirubin or transaminase levels elevated above the normal range (7, 8, 10), although precise rules have not been established. We detected greater toxicity of both agents with higher levels of bilirubin (docetaxel) or alkaline phosphatase (docetaxel and irinotecan), even for values within the normal range. The reason for greater TI in patients with normal or minimally elevated bilirubin is uncertain but could reflect alteration of pharmacokinetics response, pharmacodynamics, or both effects. Slight elevations of serum bilirubin and transaminase levels can signal underlying impairment of hepatic function sufficient to reduce drug metabolism and clearance. Alternatively, higher bilirubin or transaminase levels within the normal range could be markers of genetic variation, with reduced hepatic metabolic or excretory capacity leading to greater systemic exposure (AUC) and toxicity for a given dose. Supporting this possibility are the data of Yamamoto *et al.* (5), who found independent correlation of docetaxel clearance with cytochrome P4503A4 activity, serum levels of α-1 acid glycoprotein, and baseline transaminase activity in patients with normal total bilirubin and hepatic transaminase levels. Like docetaxel, elevated bilirubin or transaminase levels correlate with reduced clearance of paclitaxel and paclitaxel metabolites, increased AUC, and greater toxic effects (11). However, unlike docetaxel, alkaline phosphatase, total bilirubin, and transaminase values within the normal range failed to predict toxicity of paclitaxel in our population.

Phase I studies assume that dose is the significant determinant of toxicity. Our analysis of multiple Phase I and early Phase II trials reveals that dose is not always a significant predictor of toxicity. Even with conventional patient selection criteria that included the requirement for normal or near-normal hepatic and renal function, patient characteristics had greater

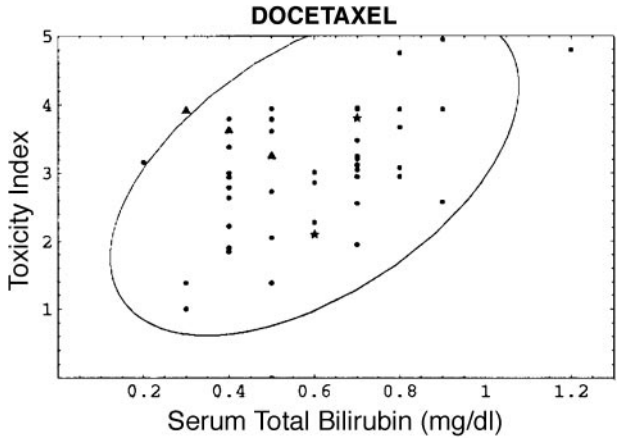

Fig. 2 Toxicity index (*Y* axis) by serum total bilirubin (mg/dl, *X* axis) in patients treated with docetaxel. *Dots* are observed values for IRB95111, the *ellipse* encloses 95% of the distribution, and the Pearson correlation coefficient is $r = 0.526$ with $P = 0.0003$. IRB99018 (★) and IRB99024 (▲) were not included either in the construction of the ellipse or the correlation coefficient because of the small sample size.

predictive value than dose for the toxicity for several agents (Table 5). Thus, the current eligibility criteria for most Phase I and II trials do not provide populations that are uniform enough to conclude that differences in toxic response are primarily dose-related. Because selecting more homogenous populations is not practical, individual variability should be considered by using statistical designs other than the typical "3+3" Phase I algorithm, which fails to account for covariates.

None of the methods used for calculating drug dose [by body surface area (mg/m^2), by weight (mg/kg), or uncorrected for weight or surface area (fixed mg dose)] was consistently the best toxicity predictor for all selected agents. This is consistent with the lack of correlation of pharmacokinetic parameters of most cytotoxic agents with body surface area or other measures of body size (12). The observation that the TI and the occurrence of DLT both correlate poorly with the method of dose calculation further indicts the reflexive use of body surface area-based dosing of new agents in early clinical trials.

A retrospective analysis such as ours has several limitations. These include variability in the criteria used to grade toxicity and definition of DLT and lack of a validation group. This underscores the preliminary nature of this analysis and its intended use as a tool to generate hypotheses. Also, the results of this analysis may not apply to later cycles of therapy and the emergence of late or cumulative toxicities (*e.g.*, neurotoxicity) with continued treatment. However, the same methodology may be applied to investigate predictors of chronic or multiple cycle toxicity. Another limitation of our analysis is the relatively small numbers of patients studied for several of the agents. Other potential predictors of toxic response may emerge as we apply our analysis to a larger base of patients.

In summary, we have found that patient characteristics compete with dose as predictors of the toxic response to several chemotherapeutic agents. Furthermore, pretreatment characteristics, even within the normal range, may be predictors of treatment toxicity. We propose a new summary measure, the TI, for identifying predictors of toxicity. Because of its greater dynamic range, the TI may prove to be a more sensitive tool than DLT alone. Toward the goal of individualizing therapy for cancer patients, prospective studies to validate the covariates identified by the TI are under way.

ACKNOWLEDGMENTS

We are grateful for the comments and suggestions provided by Drs. Roger Cohen, John A. Ridge, and Louis Weiner. We are indebted to Gail Raisen and Mary Beard for data extraction.

REFERENCES

1. Decoster G, Stein G, Holdener EE. Responses and toxic deaths in Phase I clinical trials. Ann Oncol 1990;1:175–81.

2. Dillman RO, Koziol JA. Phase I cancer trials: limitations and implications. Mol Biother 1992;4:117–121.

3. Smith TL, Lee JJ, Kantarjian HM, Legha SS, Raber MN. Design and results of Phase I cancer clinical trials: three-year experience at M. D. Anderson Cancer Center. J Clin Oncol 1996;14:287–95.

4. Eisenhauer EA, O'Dwyer PJ, Christian M, Humphrey JS. Phase I clinical trial design in cancer drug development. J Clin Oncol 2000;18: 684–92.

5. Yamamoto N, Tamura T, Fukuoka M, Saijo N. Survival and prognostic factors in lung cancer patients treated in Phase I trials: Japanese experience. Int J Oncol 1999;15:737–41.

6. Bachelot T, Ray-Coquard I, Catimel G, et al. Multivariable analysis of prognostic factors for toxicity and survival for patients enrolled in Phase I clinical trials. Ann Oncol 2000;11:151–6.

7. Alexandre J, Bleuzen P, Bonneterre J, et al. Factors predicting for efficacy and safety of docetaxel in a compassionate-use cohort of 825 heavily pretreated advanced breast cancer patients. J Clin Oncol 2000; 18:562–73.

8. Bruno R, Riva A, Hille D, Lebecq A, Thomas L. Pharmacokinetic and pharmacodynamic properties of docetaxel: results of Phase I and Phase II trials. Am J Health Syst Pharm 1997;54:516–9.

9. Chabot GG, Abigerges D, Catimel G, et al. Population pharmacokinetics and pharmacodynamics of irinotecan (CPT-11) and active metabolite SN-38 during Phase I trials. Ann Oncol 1995;6:141–51.

10. van Groeningen CJ, Van der Vijgh WJ, Baars JJ, et al. Altered pharmacokinetics and metabolism of CPT-11 in liver dysfunction: a need for guidelines. Clin Cancer Res 2000;6:1342–6.

11. Venook AP, Egorin MJ, Rosner GL, et al. Phase I and pharmacokinetic trial of paclitaxel in patients with hepatic dysfunction: Cancer and Leukemia Group B 9264. J Clin Oncol 1998;16:1811–9.

12. Baker SD, Verweij J, Rowinsky EK, et al. Role of body surface area in dosing of investigational anticancer agents in adults, 1991–2001. J Natl Cancer Inst (Bethesda) 2002;94:1883–8.