PHARMACEUTICAL STATISTICS
Pharmaceut. Statist. 2005; **4**: 161–171
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/pst.172

Gaining acceptability for the Bayesian decision-theoretic approach in dose-escalation studies

MAIN PAPER

Yinghui Zhou*,† and Maria Lucini
Medical and Pharmaceutical Statistics Research Unit, University of Reading, Reading RG6 6FN, UK

There has recently been increasing demand for better designs to conduct first-into-man dose-escalation studies more efficiently, more accurately and more quickly. The authors look into the Bayesian decision-theoretic approach and use simulation as a tool to investigate the impact of compromises with conventional practice that might make the procedures more acceptable for implementation. Copyright © 2005 John Wiley & Sons, Ltd.

Keywords: ***first-into-man; optimal design; Bayesian decision-theoretic approach; dose-level skipping***

1. INTRODUCTION

Pharmaceutical and biotechnology companies spend a great amount of resources every year on conducting first-into-man studies, in which a new compound is administered to human subjects for the first time. Some of these compounds will go through to later phases, but others will fail due to safety concerns. Even those that pass phase I may still prove unsuccessful later, due to undetected safety problems or lack of efficacy.

*Correspondence to: Yinghui Zhou, MPS Research Unit, University of Reading, Reading, RG6 6FN, U.K.
†E-mail: y.zhou@reading.ac.uk

Zhou and Whitehead [1] describe a Bayesian decision-theoretic method for phase I dose-finding studies. The method allows investigators to learn about the quantitative relationship between doses and toxicities as the study proceeds, while maintaining safety. In this paper, we show that this optimal method is flexible and can achieve dose-finding accurately, efficiently and quickly with or without additional constraints on escalation through the available dose levels.

The main therapeutic area considered in this paper is cancer, although the principles of the method have been extended to other settings, such as healthy volunteer studies [2,3]. From a first-into-man cancer study we can either identify the maximum tolerated dose (MTD) for later phase studies, or characterize the relationship between

Pharmaceutical STATISTICS

dose and risk, or both. Usually, evidence of benefit is not assessed in such studies as they are too short-term for it to become apparent.

A World Health Organization grading on a scale from 0 to 4 is used to reflect the degree of toxicity in cancer trials. However, a simpler binary response, such as toxic or non-toxic, is commonly used in statistical methods described in the literature. For example, a WHO grade greater than 2 might be equated with 'toxicity'. In this paper, we use the simple binary response. The target dose we are interested in is defined as the TD100π: the dose at which probability of toxicity is π.

In a typical dose-escalation study, successive small groups of subjects known as *cohorts* receive doses simultaneously, and are assessed before the next cohort is dosed. Each individual receives a dose from a predefined discrete dose range $d_1 < d_2 < \cdots < d_k$. The traditional dose-escalation procedure is the so-called '3+3' design [4] which proceeds as follows. The first three subjects are given dose d_1. If there are no toxic responses, then the next three subjects receive dose d_2. If one toxic response is observed, then three additional subjects are given d_1. If none of these show toxicity, then the dose is escalated to d_2 for the third cohort. In all other cases, the trial is stopped. These escalation rules then apply at any future dose. The dose below that causing stopping due to excess toxicity is declared to be the MTD. This design is simple and easily implemented, but it is likely to result in a small sample size because of its stopping rule, and it can lead to poor accuracy of the identified MTD. Statistical evaluation has shown that the dose identified as the MTD from the '3+3' design tends to be associated with a probability of toxicity between 10% and 29% [5,6].

O'Quigley *et al.* [7] introduced a statistical approach known as the continuous reassessment method (CRM), but it is rarely implemented in oncology in its original form due to safety concerns. Various authors have proposed modified versions of the CRM [8–13]. The most common practical modifications are not to allow dose-level skipping and to start dose escalation from the lowest dose. Simulations have shown that the modified CRM can limit the risk of overdosing subjects at high toxic dose levels, although the original CRM produces more accurate estimates. Some authors have suggested using rules to stop once enough information has been gathered from a trial [14,15]. O'Quigley and Shen [16] propose a two-stage non-Bayesian method, the likelihood CRM (CRML). In this version, the probability of toxicity is updated by its maximum likelihood estimate, rather than the Bayes estimate as in the original CRM. All of these efforts are intended to help clinicians to implement the method in practice. The CRM is not assessed in this paper, as the Bayesian decision-theoretic method [1] has already been shown to be more accurate and safe in Zhou and Whitehead [17].

2. THE BAYESIAN DECISION-THEORETIC APPROACH

Since 1995, the Bayesian decision-theoretic approach [18–20] has been developed to guide clinicians to select suitable starting doses when there is little information available and then use decision theory to decide which doses to give the next subject or cohort of subjects. The relationship between the probability of toxicity and dose is expressed as the logistic regression model:

$$\log\left\{\frac{p(d)}{1-p(d)}\right\} = \theta_1 + \theta_2 \log d \qquad (1)$$

where $p(d)$ is the probability of toxicity at dose d, $\log d$ is the dose d on the logarithmic scale, θ_1 is the intercept and θ_2 is the slope of the logistic regression. The last two are unknown parameters.

Although the compound is new, clinicians may possess some 'prior' knowledge or opinion of its toxicity from similar compounds or pre-clinical studies. This knowledge or opinion can be expressed as a formal prior density. If there is little information available, then either an imposed prior or a non-informative prior can be used. In this case, the prior should be deliberately chosen to be weak in order to let data overcome its influence easily. In the Bayesian decision-theoretic approach, a conjugate prior is used: a 'pseudo-dose'

d_{-1} is specified at which we imagine that n_{-1} subjects have been treated and given t_{-1} toxicities, together with a second 'pseudo-dose' d_0 at which we suppose n_0 subjects have given t_0 toxicities; t_{-1} and t_0 need not be integers. This, mathematically, is equivalent to assuming that the probability of toxicity at d_{-1} has a beta distribution with parameters t_{-1} and $n_{-1}-t_{-1}$, and the probability of toxicity at d_0 has a beta distribution with parameters t_0 and n_0-t_0. These two independent beta distributions will form the prior density function for the unknown parameters θ_1 and θ_2, that is,

$$h_0(\theta_1, \theta_2) = \prod_{i=-1}^{0} \frac{p_i^{t_i}(1-p_i)^{n_i-t_i}}{B(t_i, n_i - t_i)} \left| \log\left(\frac{d_{-1}}{d_0}\right) \right|$$

This prior was introduced by Tsutakawa [21].

Zhou and Whitehead [1] use a pessimistic prior: at the lowest available dose, taken to be d_{-1}, $n_{-1}=3$ subjects have given $t_{-1}=0.6$ toxicities, and at the highest available dose, taken to be d_0, $n_0=3$ subjects have given $t_0=1.5$ toxicities. Thus prior opinion is that the lowest available dose is the TD20 and the highest is the TD50. This pessimistic prior usually results in the starting dose being the lowest available dose, as in the traditional '3+3' design.

A gain function is needed to make decisions on which doses to give the next subject or next cohort of subjects. Suitable gain functions are proposed in [20]. The so-called 'patient gain' gives patients the dose at which the posterior probability of toxicity is closest to the target, π. The so-called 'variance gain' seeks a dose combination that gives the smallest variance of the estimated TD100π. Haines *et al.* [22] discuss and recommend the use of the variance gain, which they referred to as 'c-optimality'.

A safety constraint [1] is used to limit risk of overdosing subjects: do not use doses at which the posterior probability of toxicity exceeds a certain value, γ ($> \pi$). For example, if the target dose is set as TD20, then doses above the TD30 might be avoided. Stopping rules [1] are used to stop a trial before the maximum number of cohorts has been treated: a 'safety stopping rule' will stop if no doses match the safety constraint, and an 'accuracy stopping rule' will stop if estimates are sufficiently accurate. The criterion for sufficient accuracy might be that the ratio of the upper to the lower limit of the 95% credibility interval of the estimated target dose is small enough, say 5.

A software package, *Bayesian ADEPT*, has been developed to implement the methodology [23]. Simulation comparisons with the CRML and the '3+3' design indicate that the Bayesian method achieves more accurate estimates of the target dose while inflicting fewer toxicities [17]. The '3+3' design is the least accurate of available methods, since trials are likely to stop too early according to its stopping rule. The CRML does not necessarily improve accuracy compared with the Bayesian method, although it tends to take larger sample sizes. Comparisons of different numbers of doses and different cohort sizes have been made [24]. It was found that choosing nine dose levels has advantages over choosing five in terms of recommending doses nearer to the true target dose. Designs with a cohort size of 1 have advantages in terms of reducing the number of subjects used, but have disadvantages in terms of accuracy.

The possibility of skipping dose levels inherent in the Bayesian decision-theoretic approach appears to make investigators reluctant to implement the method. This was one of the criticisms of the original CRM. The following simulations investigate how limits on dose-level skipping affect Bayesian dose-escalation procedures in terms of economy (number of subjects), safety (number of toxicities) and accuracy.

3. THE BACKGROUND OF THE SIMULATION STUDY

In this paper, we use as an example the trial of quercetin described by Ferry *et al.* [25], which has been used as an illustration in previous papers [17–20, 24]. Patients enrolled in the trial were suffering from a variety of forms of solid tumour and were no longer amenable to standard therapies. The response was defined as renal toxicity greater than

Pharmaceutical STATISTICS

dose and risk, or both. Usually, evidence of benefit is not assessed in such studies as they are too short-term for it to become apparent.

A World Health Organization grading on a scale from 0 to 4 is used to reflect the degree of toxicity in cancer trials. However, a simpler binary response, such as toxic or non-toxic, is commonly used in statistical methods described in the literature. For example, a WHO grade greater than 2 might be equated with 'toxicity'. In this paper, we use the simple binary response. The target dose we are interested in is defined as the TD100π: the dose at which probability of toxicity is π.

In a typical dose-escalation study, successive small groups of subjects known as *cohorts* receive doses simultaneously, and are assessed before the next cohort is dosed. Each individual receives a dose from a predefined discrete dose range $d_1 < d_2 < \cdots < d_k$. The traditional dose-escalation procedure is the so-called '3+3' design [4] which proceeds as follows. The first three subjects are given dose d_1. If there are no toxic responses, then the next three subjects receive dose d_2. If one toxic response is observed, then three additional subjects are given d_1. If none of these show toxicity, then the dose is escalated to d_2 for the third cohort. In all other cases, the trial is stopped. These escalation rules then apply at any future dose. The dose below that causing stopping due to excess toxicity is declared to be the MTD. This design is simple and easily implemented, but it is likely to result in a small sample size because of its stopping rule, and it can lead to poor accuracy of the identified MTD. Statistical evaluation has shown that the dose identified as the MTD from the '3+3' design tends to be associated with a probability of toxicity between 10% and 29% [5,6].

O'Quigley *et al.* [7] introduced a statistical approach known as the continuous reassessment method (CRM), but it is rarely implemented in oncology in its original form due to safety concerns. Various authors have proposed modified versions of the CRM [8–13]. The most common practical modifications are not to allow dose-level skipping and to start dose escalation from the lowest dose. Simulations have shown that the modified CRM can limit the risk of overdosing subjects at high toxic dose levels, although the original CRM produces more accurate estimates. Some authors have suggested using rules to stop once enough information has been gathered from a trial [14,15]. O'Quigley and Shen [16] propose a two-stage non-Bayesian method, the likelihood CRM (CRML). In this version, the probability of toxicity is updated by its maximum likelihood estimate, rather than the Bayes estimate as in the original CRM. All of these efforts are intended to help clinicians to implement the method in practice. The CRM is not assessed in this paper, as the Bayesian decision-theoretic method [1] has already been shown to be more accurate and safe in Zhou and Whitehead [17].

2. THE BAYESIAN DECISION-THEORETIC APPROACH

Since 1995, the Bayesian decision-theoretic approach [18–20] has been developed to guide clinicians to select suitable starting doses when there is little information available and then use decision theory to decide which doses to give the next subject or cohort of subjects. The relationship between the probability of toxicity and dose is expressed as the logistic regression model:

$$\log\left\{\frac{p(d)}{1-p(d)}\right\} = \theta_1 + \theta_2 \log d \qquad (1)$$

where $p(d)$ is the probability of toxicity at dose d, $\log d$ is the dose d on the logarithmic scale, θ_1 is the intercept and θ_2 is the slope of the logistic regression. The last two are unknown parameters.

Although the compound is new, clinicians may possess some 'prior' knowledge or opinion of its toxicity from similar compounds or pre-clinical studies. This knowledge or opinion can be expressed as a formal prior density. If there is little information available, then either an imposed prior or a non-informative prior can be used. In this case, the prior should be deliberately chosen to be weak in order to let data overcome its influence easily. In the Bayesian decision-theoretic approach, a conjugate prior is used: a 'pseudo-dose'

Pharmaceutical STATISTICS

or equal to WHO grade 2. There were nine doses available: 60, 120, 200, 300, 420, 630, 945, 1400, 1700 mg/m^2. In the real study, most cohort sizes were 3, and 52 patients were treated. We set the cohort size at 3 in this paper. The maximum number of patients is set at 60. We investigate designs with all nine original doses available as well as just the five doses 60, 200, 420, 945 and 1700 mg/m^2, as in [24]. The TD20 is sought and the variance gain is adopted. The prior is chosen following the pessimistic criteria outlined in [17,24]: at dose 60, 3 subjects have given 0.6 toxicities, and at dose 1700 there are 1.5 toxicities out of 3 subjects. The first cohort will receive the lowest available dose of 60 mg/m^2. A safety constraint is considered, avoiding any dose with a current subjective toxicity probability exceeding 0.3. An accuracy stopping rule is also considered, which leads to stopping when the ratio R of the upper to the lower limit of the 95% credibility interval for the TD20 falls below 5. Six scenarios are listed in Table I and presented in Figure 1: these were also used in [17,24] and were selected to represent contrasting situations.

We investigate three ways to escalate doses: SK∞ is 'no dose skipping limit', which means that dose escalation will be guided by the recommendations of the Bayesian decision-theoretic procedure; SK1 is 'skip at most one dose level'; and SK0 is 'never skip a dose level while escalating'. Under both of the rules SK0 and SK1, all subjects in the first cohort must receive the lowest available dose. Suppose that, part way through the study, the highest dose any previous patient has received is d_2. Under rule SK∞ doses assigned to the next cohort could be any on the schedule. However, under rule SK1 the doses assigned to the next cohort must be no greater than d_4, while under rule SK0 they must be no greater than d_3.

In total, we investigate 24 designs under the six scenarios: number of doses (2 types), escalation rules (3 types), with or without the safety constraint (2 types), and with or without the accuracy stopping rule (2 types). The accuracy is assessed in terms of the 'precision error', defined as the square root of the average over the simulations of $\{p(\hat{d}^*) - \pi\}^2$. Here $p(d)$ represents the true dose–response relationship used in the simulation, with the intercept and slope taken from the appropriate row of Table I, and $\hat{d}^*$ is the maximum likelihood estimate of the target dose. The number of doses administered that are over the true TD30 is another criterion for measuring the safety of a design.

4. SIMULATION RUNS UNDER THE PESSIMISTIC PRIOR

Here we first use a single simulation run to illustrate how dose escalations can differ under scenario 1 using the three different escalation rules introduced in Section 3. We use nine doses and the variance gain. From Table II, it can be seen that when both the safety constraint and the accuracy stopping rule are applied, rule SK∞ achieves the greatest accuracy, although more subjects are recruited and more toxicities are observed compared with SK0. Use of rule SK1 actually leads to most subjects being recruited and most doses being administered over the true TD30. When only the accuracy stopping rule is applied, rule SK1 is the most successful escalation strategy. When only

Table I. Scenarios used for simulation.

Scenario	Description	Intercept	Slope	TD20	TD30	TD50
1	Standard	−11.87	1.78	366	489	799
2	Toxic	−14.09	2.28	261	333	480
3	Safe	−16.41	2.28	724	921	1329
4	Steep	−25.21	3.77	553	641	799
5	Very toxic	−5.30	0.79	139	280	799
6	Very safe	−8.83	1.00	1743	2930	6993

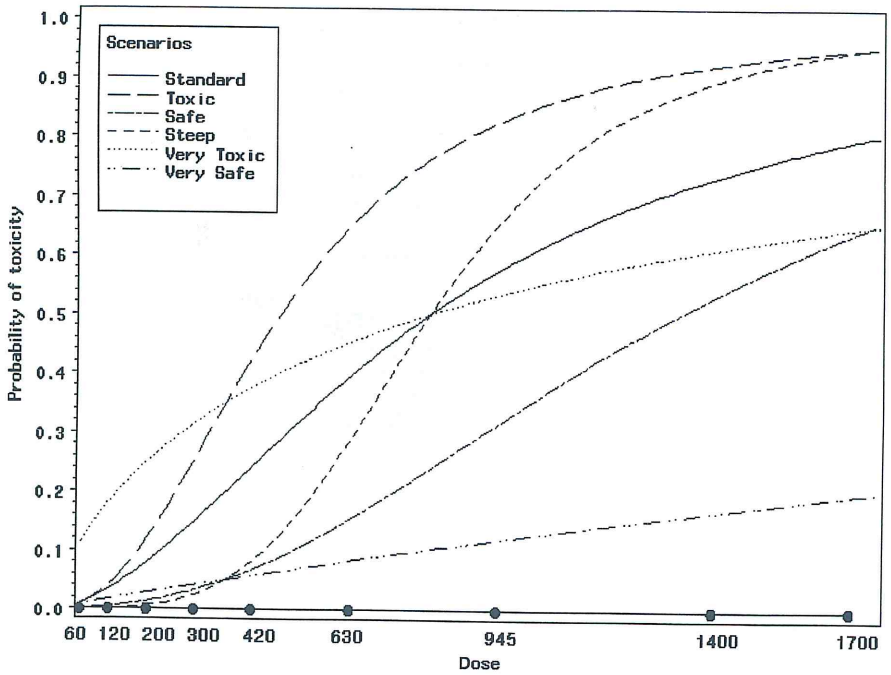

Figure 1. Summary of scenarios; the *X*-axis represents the dose levels and the *Y*-axis represents the probability of toxicity.

the safety constraint is applied, rule SK0 achieves the most accurate estimates but with an appreciable excess of doses over the true TD30 used. When neither the safety constraint nor the accuracy stopping rule is applied, rule SK1 is most successful. Figure 2 shows the progress of the four dose escalations.

We also conducted 1000 replicate simulation runs under the six different scenarios of Table I to give a clearer picture of the impact on the economy, safety and accuracy of the three dose-escalation rules. Table III summarizes the results from designs with nine doses. When both the safety constraint and the accuracy stopping rule are applied, SK∞ has advantages in terms of involving the fewest subjects, toxicities and doses over the true TD30. SK1 gives the second best results in terms of economy and safety, achieving the fewest subjects and toxicities under the 'steep' scenario. SK0 gives the fewest toxicities only under the 'very safe' scenario. When only the accuracy stopping rule is applied, both SK∞ and SK1 achieve fewer subjects and doses over the true TD30 than SK0 under the majority of scenarios. SK1 gives fewest toxicities under the 'standard' scenario and the 'steep' scenario, while SK0 gives fewest toxicities under the 'very safe' scenario. When only the safety constraint is applied, SK∞ and SK1 use fewest doses over the true TD30, while SK0 gives the fewest toxicities under most of the scenarios. When neither the safety constraint nor accuracy the stopping rule is used, SK0 gives the fewest toxicities while SK∞ and SK1 give fewest doses over the true TD30.

In practice, as discussed in Zhou and Whitehead [1], we would always recommend the use of the Bayesian approach in the presence of a safety constraint. Thus, the evaluations of SK∞ with the safety constraint are, in our view, the most relevant. They can be contrasted with applying 'never skip a dose level while escalating' as the only safety constraint, as shown in the evaluations

Pharmaceutical STATISTICS

Table II. Summary of a single simulation run under the pessimistic prior and the standard scenario (TD20 = 366).
(a) Both the safety constraint and accuracy stopping rule are applied

	No. of subjects	No. of toxicities	No. of doses > TD30	Bayes TD20 (precision error)	MLE TD20 (precision error)	95% credibility interval
SK∞	36	7	12	380 (0.011)	436 (0.055)	(177, 815)
SK1	39	7	16	467 (0.080)	531 (0.127)	(223, 984)
SK0	27	4	12	513 (0.113)	630 (0.196)	(223, 1128)

(b) Only the accuracy stopping rule is applied

	No. of subjects	No. of toxicities	No. of doses > TD30	Bayes TD20 (precision error)	MLE TD20 (precision error)	95% credibility interval
SK∞	48	9	17	443 (0.060)	493 (0.098)	(212, 926)
SK1	24	4	6	433 (0.052)	511 (0.111)	(195, 958)
SK0	33	6	3	261 (0.079)	276 (0.068)	(123, 558)

(c) Only the safety constraint applied is applied

	No. of subjects	No. of toxicities	No. of doses > TD30	Bayes TD20 (precision error)	MLE TD20 (precision error)	95% credibility interval
SK∞	60	13	2	234 (0.099)	257 (0.082)	(136, 403)
SK1	60	12	3	254 (0.084)	272 (0.071)	(155, 416)
SK0	60	11	10	353 (0.010)	369 (0.003)	(199, 626)

(d) Neither the safety constraint nor the accuracy stopping rule is applied.

	No. of subjects	No. of toxicities	No. of doses > TD30	Bayes TD20 (precision error)	MLE TD20 (precision error)	95% credibility interval
SK∞	60	12	19	427 (0.047)	463 (0.075)	(238, 767)
SK1	60	11	13	425 (0.046)	445 (0.062)	(264, 683)
SK0	60	14	16	274 (0.070)	309 (0.043)	(127, 592)

of SK0 in Table III. Such a comparison shows SK∞ with the safety constraint to be preferable to SK0 without in terms of both safety and accuracy.

Results from designs with five doses, not given here, show that it does not matter how the dose-escalation procedure is conducted if the safety constraint and the accuracy stopping rule are applied. There are no substantial differences regarding the number of subjects, toxicities and doses over the true TD30 used. This is because it is unlikely that doses will be skipped if the number of doses is small. When neither the safety constraint nor the accuracy stopping rule is used, SK0 gives fewer toxicities per run than the other rules. All of the escalation rules give similar accuracy if the safety constraint is applied. If neither the safety constraint nor the accuracy stopping rule is used, SK0 gives more accurate results than the other rules.

5. MULTIPLE SIMULATION RESULTS UNDER A REALISTIC PRIOR

The simulations described in Section 4 incorporate both the explicit safety constraint of avoiding

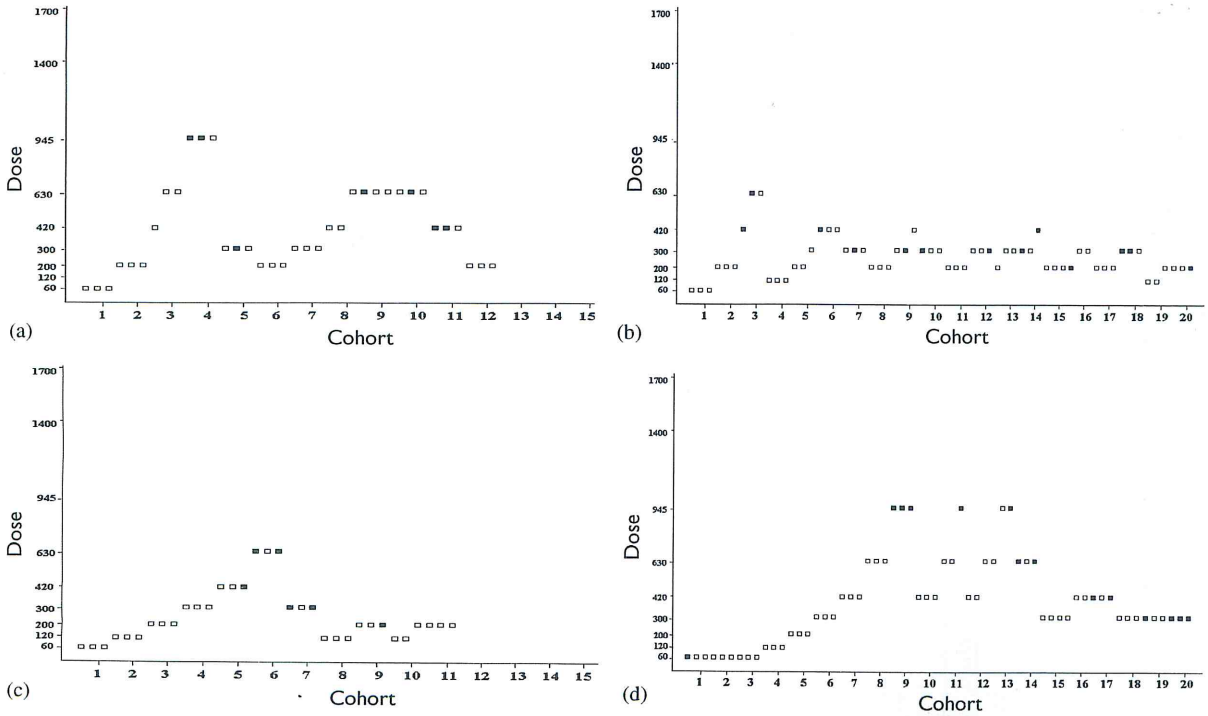

Figure 2. Progress of the simulated dose escalations analysed in Table II: (a) SK∞ with the safety constraint and the accuracy stopping rule; (b) SK∞ with the safety constraint only; (c) SK0 with the accuracy stopping rule only; (d) SK0 without the safety constraint and the accuracy stopping rule. *Open squares represent subjects who had no toxicity; solid squares represent subjects who had toxicity.*

doses currently believed to lie above the TD30 and the implicit caution of a pessimistic prior. Now, the relative performances of the dose-escalation rules are explored in the case of a realistic and quite optimistic prior. Such a prior will allow SK∞ to begin at a dose other than the lowest available. We have conducted simulations under the realistic prior: at dose 200, 3 subjects have given 0.6 toxicities, and at dose 1700 there are 1.5 toxicities out of 3 subjects. Thus 200 is the prior TD20 and 1700 the prior TD50.

The results in Table IV show that SK∞ usually uses fewest subjects per run while SK0 usually uses fewest doses over the true TD30 per run. SK∞ with the safety constraint gives fewer toxicities and uses fewer doses over the true TD30 per run than SK0 without the safety constraint when the stopping rule is not applied. When the accuracy stopping rule is applied, SK∞ has clear advantages in terms of number of subjects and accuracy, while SK0 uses fewer doses over the true TD30. However, there is no winner in terms of toxicities. If the safety is of greater concern, then we would suggest use the SK∞ design with the safety constraint and with the accuracy stopping rule, even though it will be less accurate. If the accuracy is of greater concern, then we would suggest use the SK∞ design with the safety constraint but without the accuracy stopping rule, despite its greater number of toxicities.

6. DISCUSSION

From the multiple simulations under the pessimistic prior, results indicate that when the escalation rule SK∞ is used with the safety constraint and

Table III. Results from all of the designs under the pessimistic prior. The best result for each scenario in each column is shown in bold.

		Stopping rules applied															
		Safety and accuracy				Accuracy only				Safety only				Neither			
Scenario	Escalation rule	No. of subjects per run	No. of toxicities per run	No. of doses > true TD30	Precision error	No. of subjects per run	No. of toxicities per run	No. of doses > true TD30	Precision error	No. of subjects per run	No. of toxicities per run	No. of doses > true TD30	Precision error	No. of subjects per run	No. of toxicities per run	No. of doses > true TD30	Precision error
Standard	SK∞	**36.6**	**6.6**	7.1	0.095	35.4	6.6	6.7	0.099	60	11.4	9.9	**0.055**	60	11.7	10.2	0.059
	SK1	36.7	6.7	**6.2**	**0.089**	**35.3**	**6.5**	**5.9**	**0.097**	60	11.4	**9.0**	0.056	60	11.7	**9.4**	0.058
	SK0	37.0	6.7	6.4	0.099	36.5	6.7	6.6	0.103	60	**11.3**	9.1	0.060	60	**11.6**	9.9	**0.056**
Toxic	SK∞	**35.7**	**7.0**	**6.8**	0.100	**34.6**	**6.9**	**6.5**	0.110	60	12.0	**9.9**	0.058	60	12.1	**10.0**	0.060
	SK1	36.0	**7.0**	7.5	**0.095**	35.1	**6.9**	7.3	**0.010**	60	**11.9**	10.6	**0.053**	60	12.1	10.7	**0.059**
	SK0	36.5	7.1	7.5	0.106	35.8	7.0	7.4	0.108	60	**11.9**	10.6	0.060	60	12.1	10.8	**0.059**
Safe	SK∞	**31.4**	5.0	**8.2**	**0.106**	**29.5**	**4.8**	**6.5**	**0.119**	60	10.3	**16.9**	0.061	60	10.7	**15.2**	0.062
	SK1	31.7	5.0	9.0	0.120	29.8	4.9	7.9	0.118	60	10.3	17.4	0.061	60	10.7	16.4	**0.060**
	SK0	31.7	5.0	9.0	0.120	30.7	4.9	8.3	0.125	60	**10.1**	17.2	0.061	60	**10.4**	16.4	0.061
Steep	SK∞	27.2	4.5	**2.6**	0.170	**25.4**	4.5	**2.6**	0.175	60	10.8	**3.4**	0.067	60	11.5	**4.3**	0.067
	SK1	**26.2**	**4.3**	3.8	0.174	**25.4**	**4.3**	3.8	**0.173**	60	10.9	4.4	**0.063**	60	11.2	5.7	**0.066**
	SK0	27.3	4.5	3.3	**0.158**	26.3	4.5	3.6	**0.173**	60	**10.7**	4.1	0.065	60	**11.1**	5.5	**0.066**
Very toxic	SK∞	**53.2**	**10.7**	10.0	0.067	**52.9**	11.1	10.2	0.066	60	12.3	11.4	0.058	60	12.7	11.7	0.057
	SK1	53.4	**10.7**	**9.8**	**0.065**	53.0	**10.9**	**10.0**	**0.065**	60	**12.2**	**11.2**	**0.057**	60	12.5	**11.5**	**0.056**
	SK0	53.8	10.8	9.9	0.068	53.6	**10.9**	10.4	0.067	60	**12.2**	11.4	0.060	60	**12.3**	11.8	0.059
Very safe	SK∞	53.3	6.8	0	**0.052**	53.4	6.9	0	**0.051**	60	7.7	0	**0.046**	60	7.8	0	**0.045**
	SK1	**52.2**	6.7	0	0.055	**52.8**	6.9	0	0.055	60	7.8	0	**0.046**	60	7.9	0	0.046
	SK0	52.9	**6.3**	0	0.056	53.5	**6.5**	0	0.056	60	**7.3**	0	0.050	60	**7.4**	0	0.049

Pharmaceutical STATISTICS

Table IV. Results from all of the designs under the realistic prior. The best result for each scenario in each column is shown in bold.

		Stopping rules applied															
		Safety and accuracy				Accuracy only				Safety only				Neither			
Scenario	Escalation rule	No. of subjects per run	No. of toxicities per run	No. of doses > true TD30	Precision error	No. of subjects per run	No. of toxicities per run	No. of doses > true TD30	Precision error	No. of subjects per run	No. of toxicities per run	No. of doses > true TD30	Precision error	No. of subjects per run	No of toxicities per run	No. of doses > true TD30	Precision error
Standard	SK∞	**26.3**	5.0	**4.7**	**0.113**	**25.1**	**4.8**	**4.6**	**0.115**	60	12.0	9.0	**0.054**	60	12.3	10.0	0.055
	SK1	26.5	5.0	4.9	**0.113**	25.6	5.0	4.8	0.117	60	**11.9**	**8.9**	0.056	60	12.3	**10.2**	**0.054**
	SK0	26.9	5.0	4.8	0.117	26.6	5.1	4.8	0.117	60	**11.9**	9.4	0.059	60	**12.2**	10.4	0.058
Toxic	SK∞	**28.3**	5.8	6.6	0.118	**27.6**	**5.7**	6.4	0.117	60	12.5	11.2	0.059	60	12.7	11.7	0.059
	SK1	**28.3**	5.8	6.1	**0.116**	27.9	**5.7**	**5.9**	**0.116**	60	**12.4**	**10.3**	0.058	60	12.7	11.0	0.058
	SK0	29.0	**5.0**	**4.8**	0.117	28.5	5.8	6.1	0.122	60	12.5	10.5	**0.057**	60	**12.6**	**10.9**	**0.056**
Safe	SK∞	**20.6**	3.3	5.8	0.146	**20.0**	3.4	5.6	0.146	60	10.7	17.3	0.061	60	11.4	18.1	0.059
	SK1	**20.6**	3.3	5.8	0.146	20.2	3.4	5.6	0.145	60	10.7	17.4	0.061	60	11.4	**18.0**	0.058
	SK0	21.3	3.3	**5.4**	**0.139**	20.8	**3.3**	**5.5**	**0.139**	60	10.7	**16.7**	**0.052**	60	**11.3**	**18.0**	**0.056**
Steep	SK∞	**17.7**	**3.2**	2.4	0.194	**17.4**	**3.2**	2.6	0.189	60	11.6	3.1	0.063	60	12.5	4.8	0.063
	SK1	**17.7**	**3.2**	2.4	0.195	**17.4**	**3.2**	2.6	**0.186**	60	11.6	3.1	0.063	60	12.2	4.8	0.063
	SK0	19.0	3.3	**2.0**	**0.190**	18.4	3.4	**2.4**	0.195	60	11.6	**2.9**	0.063	60	**12.2**	**4.6**	0.063
Very toxic	SK∞	49.1	10.0	9.7	0.079	**47.6**	10.1	10.2	0.087	60	12.5	12.7	**0.058**	60	13.0	13.6	0.058
	SK1	**48.9**	**9.9**	10.0	0.082	47.7	**9.9**	10.2	0.085	60	**12.4**	12.5	0.059	60	**12.7**	13.5	0.058
	SK0	**48.9**	**9.9**	**9.2**	**0.079**	48.6	**9.9**	**9.6**	**0.082**	60	**12.4**	**12.3**	0.065	60	12.9	**12.9**	0.062
Very safe	SK∞	40.0	5.0	0	0.062	40.5	5.1	0	**0.062**	60	7.6	0	**0.048**	60	7.8	0	0.046
	SK1	40.2	4.9	0	**0.061**	40.6	5.1	0	**0.062**	60	7.6	0	0.049	60	7.8	0	**0.045**
	SK0	**38.3**	**4.5**	0	0.066	**40.2**	**4.8**	0	0.064	60	7.6	0	0.051	60	7.8	0	0.049

Pharmaceutical STATISTICS

the accuracy stopping rule, it has a clear advantage in terms of overdosing control (judged by the number of doses over the true TD30) and accuracy, compared with the escalation rule SK0 with the accuracy stopping rule only. When SK∞ is used with the safety constraint only, it reduces the number of toxicities and doses over the true TD30 and improves accuracy, compared with SK0 without the safety constraint. From the multiple simulations under the realistic prior, results show that when SK∞ is used with the safety constraint and the accuracy stopping rule, it reduces the number of subjects and improves accuracy, compared with SK0 with the accuracy stopping rule only. When SK∞ is used with the safety constraint only, it reduces the number of toxicities and doses over the true TD30, compared with SK0 without the safety constraint. We hope that these results will give clinicians more confidence in using the Bayesian decision-theoretic approach. This approach not only predicts the optimal target dose and estimates the dose–response curve, but also addresses safety concerns via the safety constraint.

In this paper we have compared escalation strategies while maintaining a constant number of available doses. However, the Bayesian approach can be used with large numbers of available doses without lengthening the duration of the study. To see this, suppose that initially nine doses are proposed and a conventional escalation strategy using SK0 envisaged. If now all mid-doses are added to the schedule, 17 doses become available. A Bayesian strategy with SK1 would allow dosing at the most appropriate levels, while being no more radical than a conventional procedure using SK0 with nine doses. In general, it should be borne in mind that all procedures do allow 'dose-level skipping'; as all doses lying in the continuous range between the chosen doses that are available will always be skipped. The Bayesian approach applied with many doses and some allowance of skipping provides a flexible solution.

ACKNOWLEDGEMENTS

The authors would like to thank John Whitehead for his useful comments, and the referees for helpful suggestions.

REFERENCES

1. Zhou Y, Whitehead J. Practical implementation of Bayesian dose-escalation procedures. *Drug Information Journal* 2003; **37**:45–59.
2. Patterson S, Francis S, Ireson M, Webber D, Whitehead J. A novel Bayesian decision procedure for early phase dose-finding studies. *Journal of Biopharmaceutical Statistics* 1999; **9**:583–597.
3. Whitehead J, Zhou Y, Patterson S, Webber D, Francis S. Easy-to-implement Bayesian methods for dose-escalation studies in healthy volunteers. *Biostatistics* 2001; **2**:47–61.
4. Carter SK. Study design principles for the clinical evaluation of new drugs as developed by the chemotherapy programme of the National Cancer Institute. In *The Design of Clinical Trials in Cancer Therapy*, Staquet MJ (ed.). Scientifiques Européennes: Brussels, 1973, pp. 242–289.
5. Lin Y, Shih WJ. Statistical properties of the traditional algorithm-based designs for phase I cancer clinical trials. *Biostatistics* 2001; **2**:203–215.
6. Kang SH, Ahn CW. An investigation of the traditional algorithm-based designs for phase I cancer clinical trials. *Drug Information Journal* 2002; **36**:865–873.
7. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* 1990; **46**:33–48.
8. Faries D. Practical modifications of the continual reassessment method for phase I cancer clinical trials. *Journal of Biopharmaceutical Statistics* 1994; **4**:147–164.
9. Korn EL, Midthune D, Chen TT, Rubinstein LV, Christian MC, Simon RM. A comparison of two phase I designs. *Statistics in Medicine* 1994; **13**:1799–1806.
10. Goodman SN, Zahurak ML, Piantadosi S. Some practical improvements in the continual reassessment method for phase I studies. *Statistics in Medicine* 1995; **14**:1149–1161.
11. Moller S. 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* 1995; **14**:911–922.
12. Ahn C. An evaluation of phase I cancer clinical trial design. *Statistics in Medicine* 1998; **17**: 1537–1549.
13. Heyd JM, Carlin BP. Adaptive design improvements in the continual reassessment method for phase I studies. *Statistics in Medicine* 1999; **18**: 1307–1321.
14. O'Quigley J, Reiner E. A stopping rule for the continual reassessment method. *Biometrika* 1998; **85**:741–748.

15. Zohar S, Chevret S. The continual reassessment method: comparison of Bayesian stopping rules for dose-ranging studies. *Statistics in Medicine* 2001; **20**:2827–2843.
16. O'Quigley J, Shen LZ. Continual reassessment method: a likelihood approach. *Biometrics* 1996; **52**:673–684.
17. Zhou Y, Whitehead J. Practical implementation of Bayesian dose-escalation procedures. *Journal of the Drug Information Association* 2003; **37**:45–59.
18. Whitehead J, Brunier H. Bayesian decision procedures for dose determining experiments. *Statistics in Medicine* 1995; **14**:885–893.
19. Whitehead J. Bayesian decision procedures with application to dose-finding studies. *International Journal of Pharmaceutical Medicine* 1997; **11**: 201–208.
20. Whitehead J, Williamson D. An evaluation of Bayesian decision procedures for dose-finding studies. *Journal of Biopharmaceutical Statistics* 1998; **8**:445–467.
21. Tsutakawa RK. Bayesian inference for bioassay. *Technical report 52*. Mathematical Sciences, University of Missouri, Columbia, 1975.
22. Haines LM, Perevozskaya I, Rosenberger WF. Bayesian optimal designs for phase I clinical trials. *Biometrics* 2003; **59**:591–600.
23. Zhou Y, Whitehead J. *Bayesian ADEPT: Operating Manual*. The University of Reading, 2002.
24. Zhou Y. Choosing the number of doses and the cohort size for phase I dose-escalation studies. *Drug Information Journal* (to appear).
25. Ferry DR, Smith A, Malkandi J, Fyfe DW, Detakats PG, Anderson D, Baker J, Kerr DJ. Phase I clinical trial of the flavonoid quercetin – pharmacokinetics and evidence for *in vivo* tyrosine kinase inhibition. *Clinical Cancer Research* 1996; **2**: 659–668.