bivariate Continuous Reassessment Method

Tessella Project Number 3760

Simulation Test Results

Tessella Support Services plc

e-mail: post@tessella.com **www.tessella.com**

Head Office: Abingdon

3 Vineyard Chambers, Abingdon, Oxfordshire OX14 3PX, England
Tel: +44 (0)1235 555511
Fax: +44 (0)1235 553301

Document Control

This document is under document control. All members of the distribution list will receive updated copies whenever alterations are made.

Contributors

Person	Role	Company	Contribution
Tom Parke	Author	Tessella	

Document Approval

Person	Role	Company

Distribution

Person	Role	Company
Notes Projects Database 3760		Tessella

Revision History

Issue	Author	Date	Description
V1.R1.M0	Tom Parke	23-Mar-2005	1st version written for release v1.0.0
V1.R2.M0	Robert Nelson	27-Jul-2005	Version for release v1.0.1
V1.R3.M0	Robert Nelson	05-Oct-2005	Version for release v1.1.2

References

Ref.	Document	Date	Details and Version
[Goodman]	"Some practical improvements in the continual reassessment method for phase I studies", S.N.Goodman, M.L.Zahurak, S.Piantadosi	Statistics in Medicine 14:1149-1161, 1995	Paper describing a practical application of the CRM to phase I studies
[Braun]	"The bivariate continual reassessment method: extending the CRM to phase I trials of two competing outcomes", T.M.Braun	Controlled Clinical Trials 23:240-256, 2002	Original paper describing the bivariate continual reassessment method

Table of Contents

1 Introduction

1.1 Purpose of this document

This document describes the results of testing bCRM application by simulation. It will be useful for those intending to use the system. Users are encouraged to read the 1995 [Goodman] paper and the original bCRM paper [Braun] before using the bCRM application.

1.2 Scope of this document

This testing primarily covers the setting up and running of simulations in bCRM.

1.3 Context of this Issue

This issue documents version 1.1.2 of the system.

1.4 Definition of Terms

bCRM	refers to the computer programs that implement the bivariate continual reassessment method dose-finding algorithm, simulate and run phase 1 clinical trials using the bivariate continual reassessment method design and provide charts and graphs of the results.
A Simulation Series	is comprised of a 'Design' for a trial using the bCRM algorithm, a set of one or more 'Scenarios' – possible probabilities of toxicities and efficacies – and one or more 'Variants' where some of the Design parameters can be varied to explore the effect of the variation on the operating characteristics of a Trial. A Simulation Series creates a number of Runs = (number of Scenarios) * (number of Variants).
A Design	is a set of parameters that define a trial design to be carried out using the bCRM algorithm. This consists of: the specific drug dose strengths that can be used during the trial; the 'alpha' value of the modelling logistic curve along with a set of estimated toxicity and efficacy probabilities which are used to define the set of curves to model the response; the start dose and maximum dose level increment allowed during the trial; and the trial size details – the number of subjects assigned in each cohort used in the trial, the maximum number of subjects and the trial termination conditions.
A Scenario	is a set of fixed values of the probability of toxicity and efficacy for each of the dose strengths to be used in the trial. These define the 'true' drug response during a simulation.
A Variant	is a small modification to a Design – for example a different cohort size, study sample size or value of alpha.

A Run	A run is a set of all the parameters required to simulate a trial, it is the combination of a Design and a Scenario. A Simulation Series creates one or more runs. Each run is organised as a separate folder, all the parameters for a run are held in a single file called 'init.bcrm' and all the outputs for the simulations of the run are held in the folder in '.csv' files. (Plain text files with one record per line, individual values separated by commas – this format of file is readily imported into many other programs such as Excel, Access, SAS and Oracle). bCRM can have at most one run directory open at any one time and it is the files in this directory that provide the source data for the parameter values and visualisation.
Simulation	A simulation is the result of probabilistically generating a single clinical trial using bCRM. bCRM creates data for subjects and their responses by randomly selecting from the relevant probability distributions of the scenario defined in the bCRM input file. A particular run may be simulated a number of times, indeed to accurately analyse the characteristics of a particular Design it must be simulated ~10,000 times over a range of scenarios.
Trial	A trial is a real clinical trial, where the user enters subjects' data and responses. The bCRM Algorithm is run to determine the doses to allocate and to analyse the results. This facility allows bCRM's adaptive allocation to be used on small (single centre) trials.
MTD	The maximum tolerable dose. This is the strongest of the doses in the trial that has a toxicity below the target toxicity.
MED	The minimum efficacious dose. This is the weakest of the doses in the trial that has an efficacy above the target efficacy.

2 Test Plan

The two papers that were the basis for the implementation of the bCRM are [Goodman] and [Braun]. Both these papers contain example trial designs and the results from simulating them. bCRM was tested by entering these trial designs and running our own simulations and comparing the results with those quoted in the papers.

[Goodman] has a test design with 6 scenarios that are simulated under a number of variants of the design (different cohort sizes and different priors). Goodman is only concerned with toxicity, so for these tests we disable the second, MED finding, stage of the algorithm.

[Braun] has two examples, the first with 7 scenarios simulated with just one design and the second (taken from a paper by Gooley at al) has 3 scenarios. The first example has 'conventional', efficacy and response curves that vary gradually with dose. The second example has almost step function response and efficacy curves with varying sized windows of doses that have the minimum efficacy and acceptable toxicity.

3 Summary of Test Results

The test results using the example from [Goodman] and good correspondence with the results reported by him, show that the underlying implementation is correct.

The test results using the examples from [Braun] show that (with the exception of a very shallow toxicity curve) there are realistic designs where the system gets the MTD and MED doses correctly roughly 50% of the time and correct to within 1 dose level 90-95% of the time.

The algorithm does a good job of avoiding exposing subjects to unwanted levels of toxicity.

Thus when used in a safety study using patients rather than healthy volunteers, the system can help in assessing whether there is likely to be a viable dose range that can be used in phase II and to within one dose level, what that dose range is.

4 Detailed Test Results

4.1 Toxicity only examples from [Goodman]

The basic design is a study with 6 doses, a study size of 18-24 subjects with a target toxicity of 20%. Early termination is allowed after 6 subjects have been allocated to MTD. [Goodman] gives results for cohorts of 1, 2 and 3 subjects. We just simulate cohorts of size 2.

In bCRM version 1.0.0, modifications had to be made to the study in order to match [Goodman]. bCRM v1.1.2 can be configured to match [Goodman].

4.1.1 Design Parameters

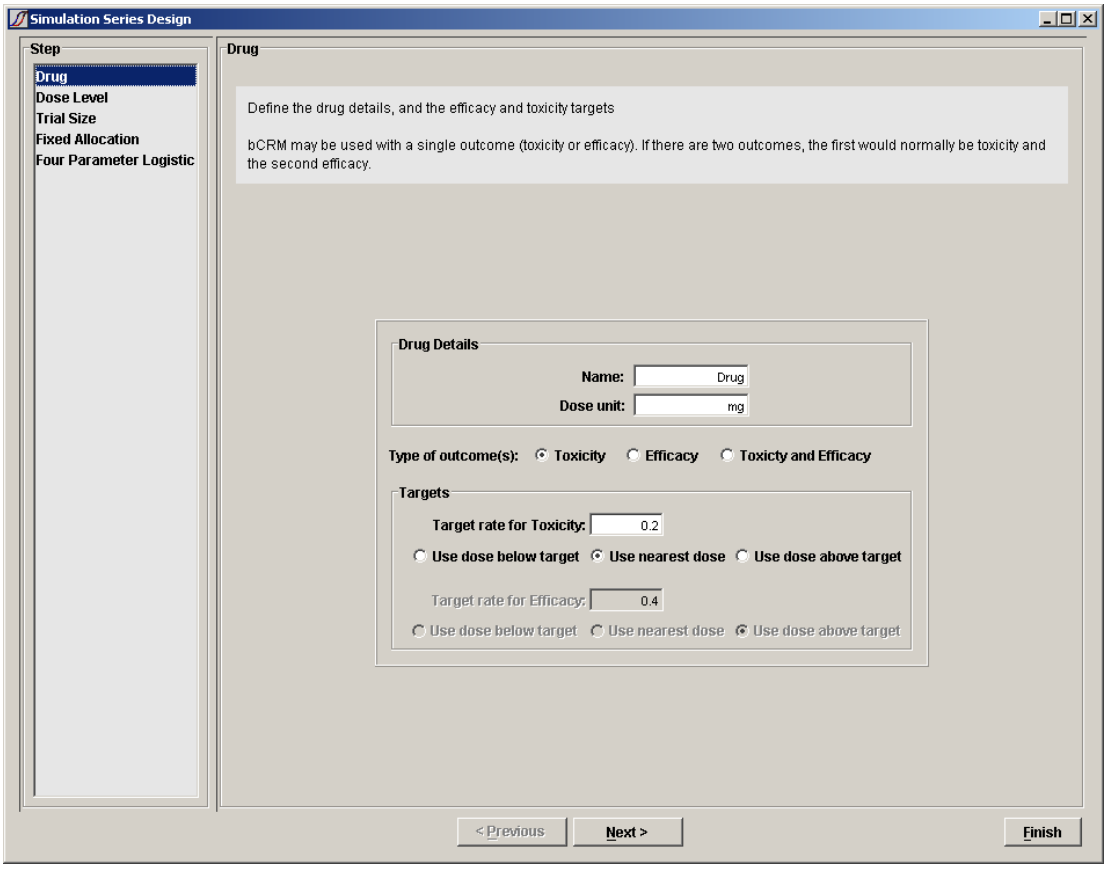


Figure 4-1

The target probability of toxicity is 20% and the target is the nearest dose, as in [Goodman].

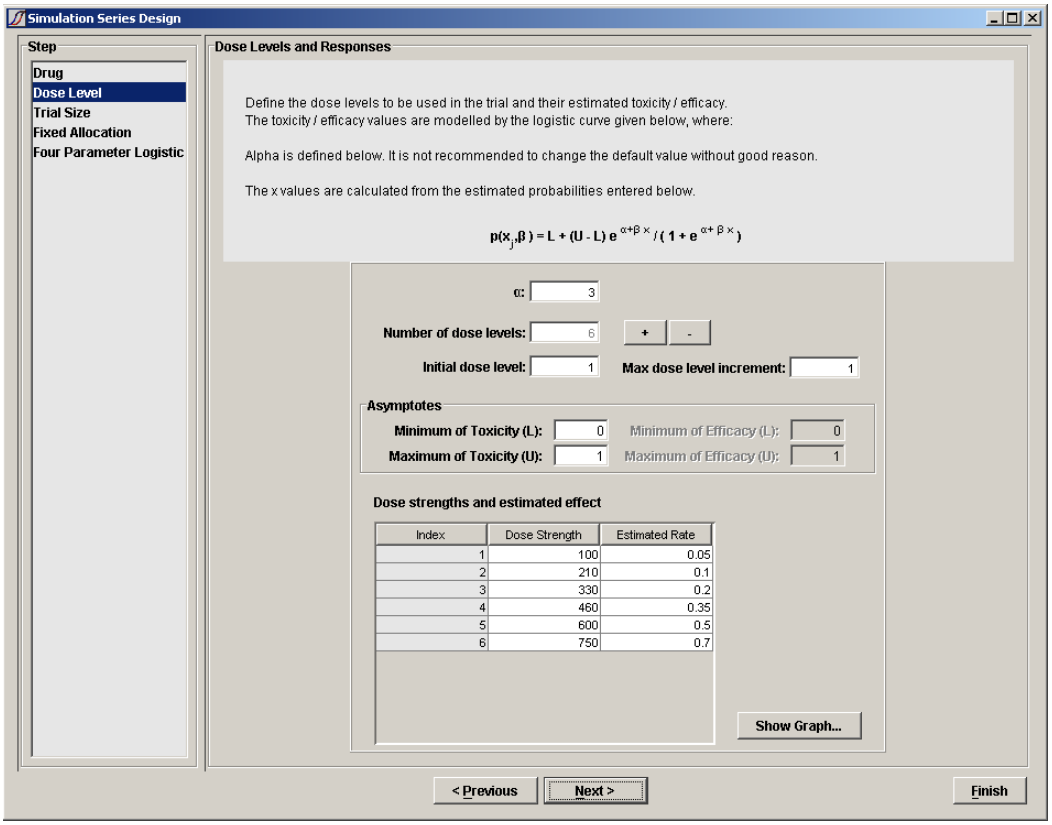


Figure 4-2

The prior toxicity estimates and alpha of +3 give a family of curve's that can represent a broad range of toxicity for all the doses. The algorithm is instructed to start at the lowest dose and increment in single dose steps (Goodman's 'Modified CRM').

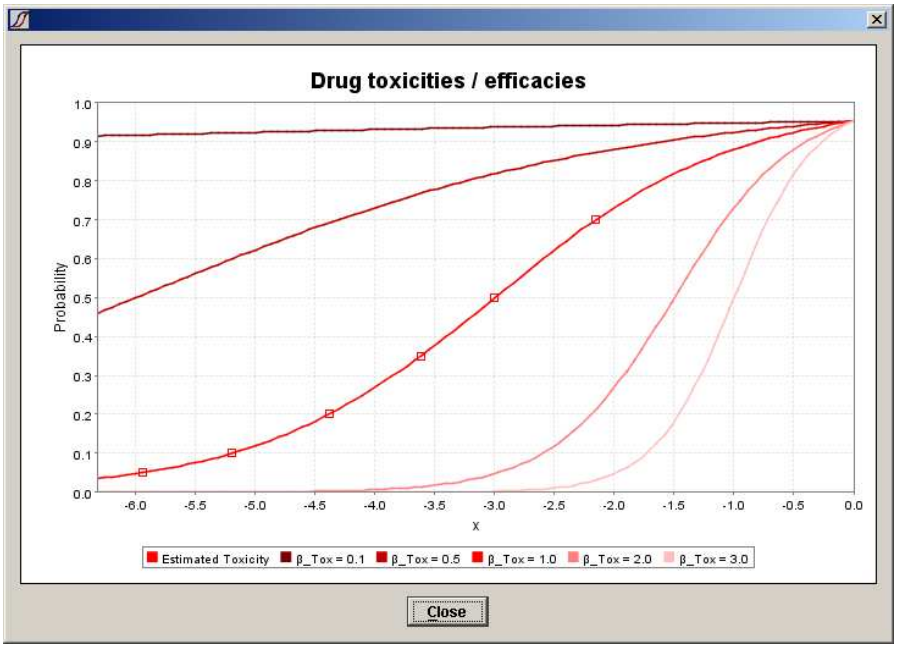


Figure 4-3

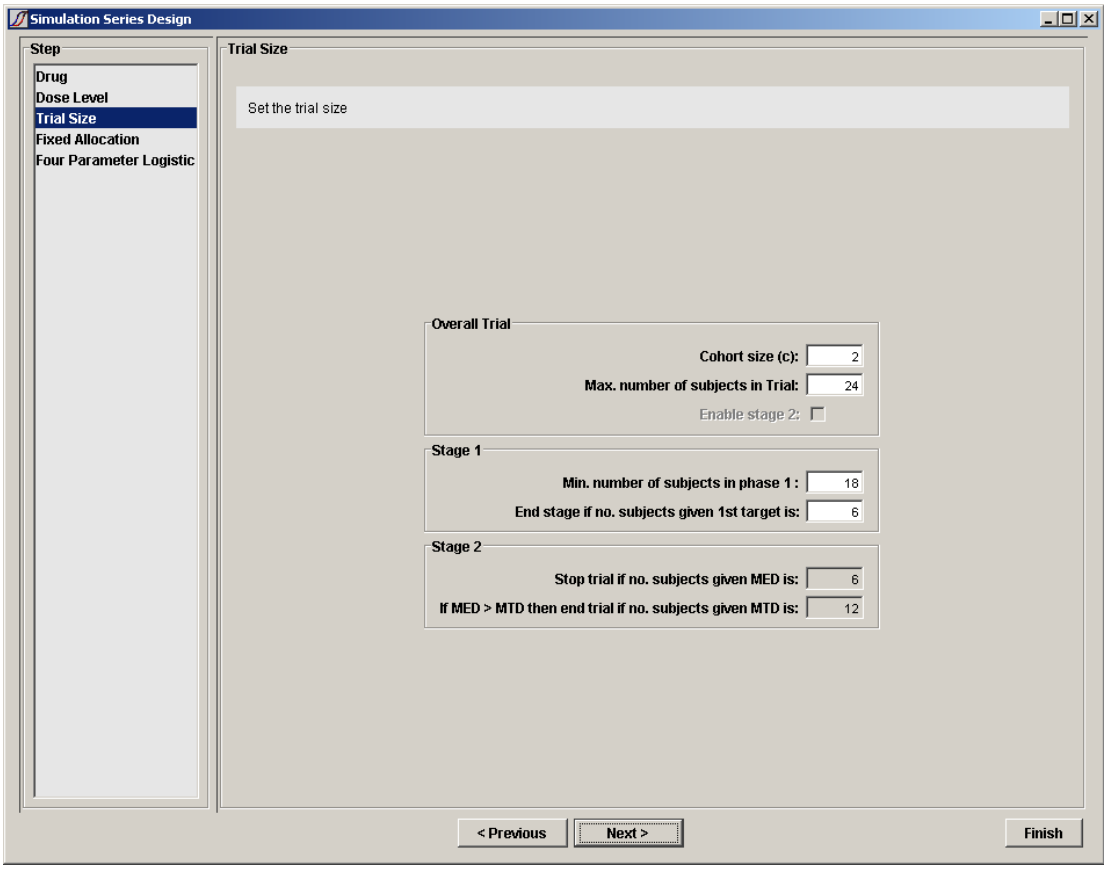


Figure 4-4

We use cohorts of size 2, with a minimum of 18 subjects and a maximum of 24.

Allocation of extra subjects to the highest and lowest doses was not enabled and neither was the fitting of a four parameter logistic.

4.1.2 Scenarios

These are exactly as in [Goodman]:

Dose Level	Curve 1 (and Prior)	Curve 2	Curve 3	Curve 4	Curve 5	Curve 6
1	0.05	0.05	0.10	0.01	0.30	0.05
2	0.10	0.10	0.10	0.01	0.40	0.05
3	0.20	0.15	0.10	0.05	0.52	0.05
4	0.35	0.20	0.10	0.10	0.61	0.05
5	0.50	0.25	0.25	0.25	0.76	0.10
6	0.70	0.35	0.80	0.80	0.87	0.15

Table 4-1

Simulated Probability of Toxicity	
Curve 1	Curve 2

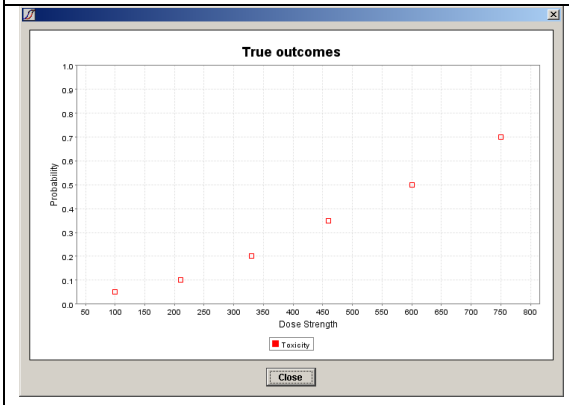
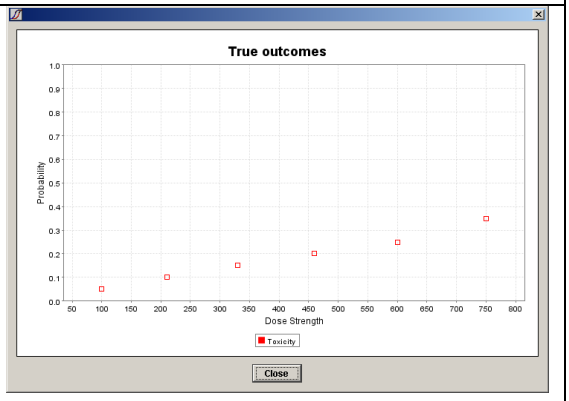

	

Figure 4-5	**Figure 4-6**
Curve 3	Curve 4

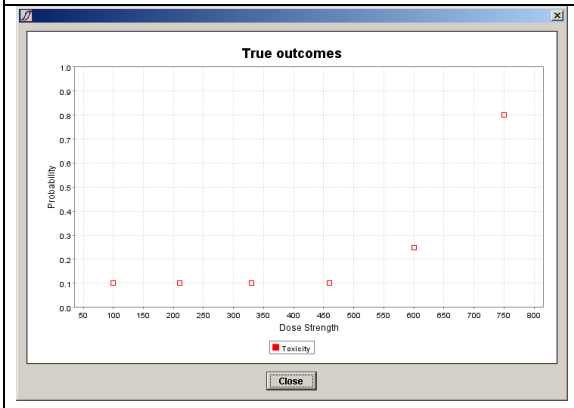
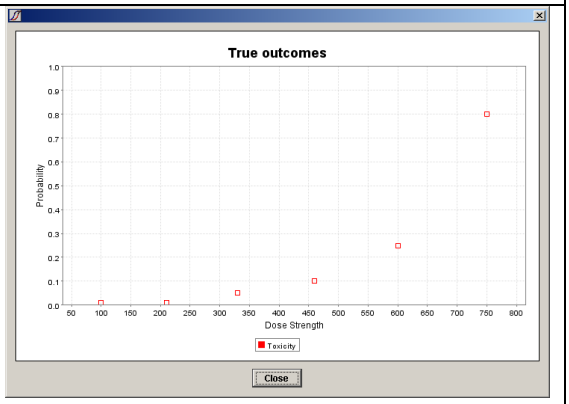

	

Figure 4-7	**Figure 4-8**
Curve 5	Curve 6

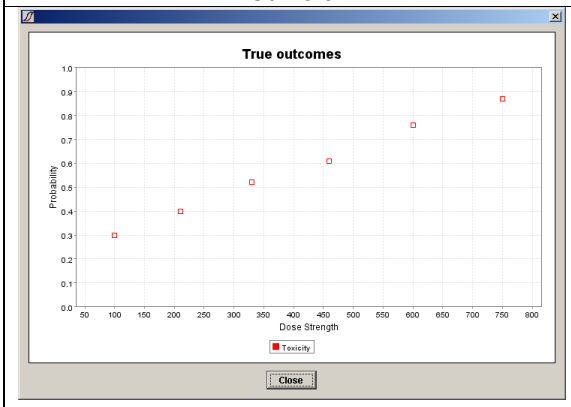
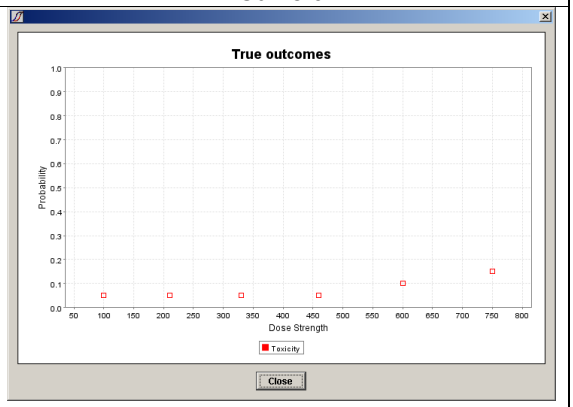

	

Figure 4-9	**Figure 4-10**

Table 4-2

4.1.3 Results

Distributions of found MTD over 1,000 simulations	
Curve 1	Curve 2

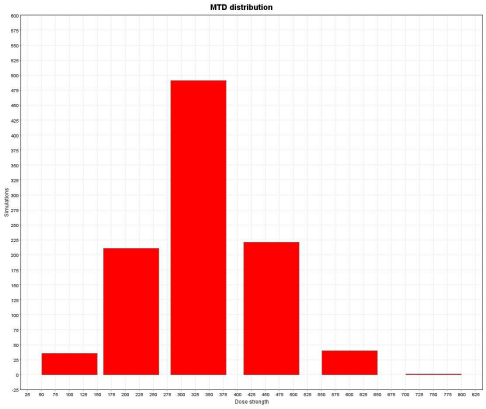
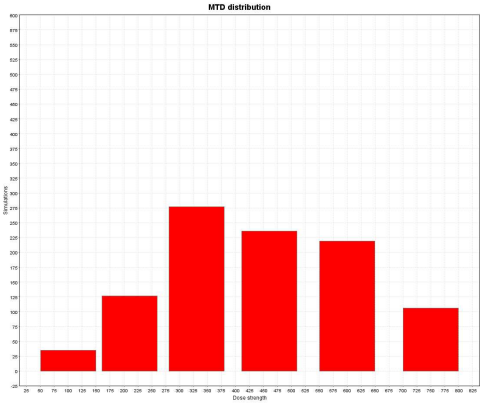

	

Figure 4-11	**Figure 4-12**
Curve 3	Curve 4

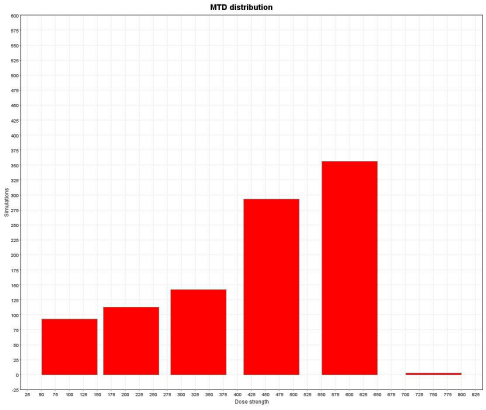
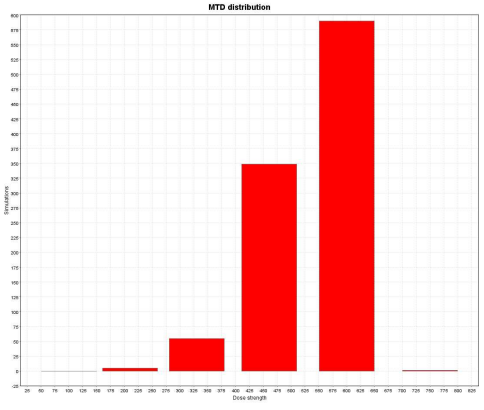

	

Figure 4-13	**Figure 4-14**
Curve 5	Curve 6

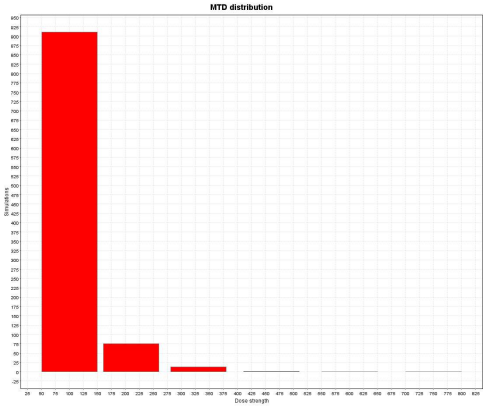
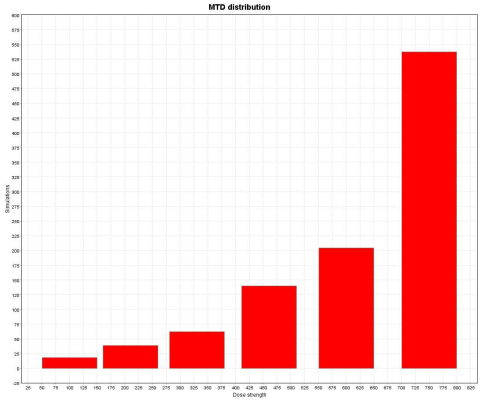

	

Figure 4-15	**Figure 4-16**

Table 4-3

These distributions are within a few percent of the results for the similar case in [Goodman] (2 subjects per cohort, modified CRM).

The following table compares the average allocation to each dose for the 6 scenarios, with Goodman's results for 2 cohorts and constant prior. The results are very similar and we are reassured that these results are sufficient similar to conclude that (disabling the extension to find the MED) we have correctly implemented the Goodman algorithm. The following table has a format based on that in [Goodman].

	% Allocation to Doses						Percent	Average
	1	**2**	**3**	**4**	**5**	**6**	**Tox**	**No. Subj**
C1 Toxicity	*5%*	*10%*	*20%*	*35%*	*50%*	*70%*		
Goodman	16	22	31	21	8	1	21.2	18.8
bCRM	19	23	31	18	5	1	22.0	18.9
C2 Toxicity	*5%*	*10%*	*15%*	*20%*	*25%*	*35%*		
Goodman	16	18	23	20	15	8	16.6	18.8
bCRM	19	19	23	18	13	7	18.7	18.8
C3 Toxicity	*10%*	*10%*	*10%*	*10%*	*25%*	*80%*		
Goodman	20	16	18	19	21	5	16.8	19.0
bCRM	25	17	17	18	18	4	18.1	18.8
C4 Toxicity	*1%*	*1%*	*5%*	*10%*	*25%*	*80%*		
Goodman	11	11	14	23	32	8	17.8	18.8
bCRM	12	12	16	25	29	7	18.7	18.6
C5 Toxicity	*30%*	*40%*	*52%*	*61%*	*76%*	*87%*		
Goodman	75	17	7	1	0	0	33.6	18.2
bCRM	81	13	5	1	0	0	34.0	18.2
C6 Toxicity	*5%*	*5%*	*5%*	*5%*	*10%*	*15%*		
Goodman	14	13	15	15	18	25	8.4	19.0
bCRM	16	14	15	15	15	24	8.0	19.1

Table 4-4

This data is included on the bCRM CD as the simulation series "Goodman2".

(The results from bCRM 1.0.1 are also included as "Goodman Example". These used a target of 0.25 but searched for an MTD which was below this value resulting in some differences from Goodman)

4.2 Toxicity and efficacy examples from [Braun]

4.2.1 Design Parameters

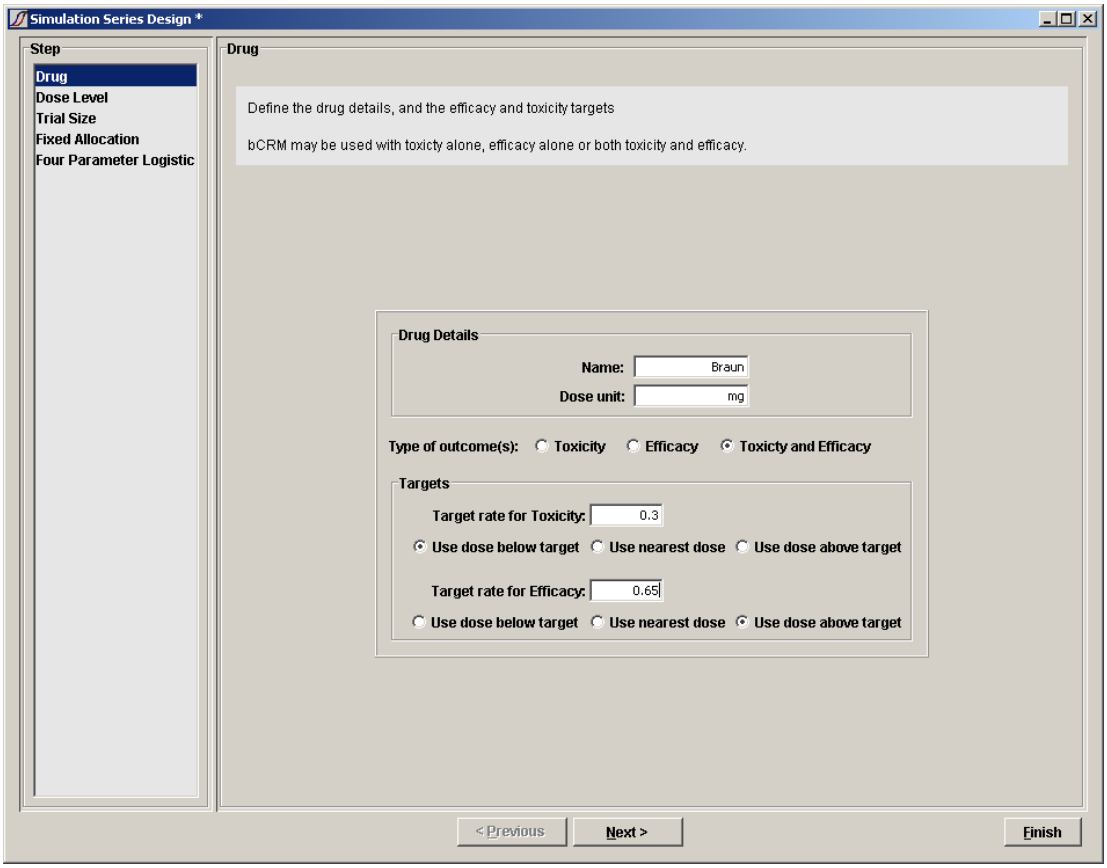


Figure 4-17

In [Braun] the targeted rates of toxicity (aGVHD) and disease progression are 25% and 30%. To allow for our algorithm treating these as limits rather than targets we use figures of 30% and 35%. Because we use efficacy, we use the rate that is the converse of the disease progression rate – 65%.

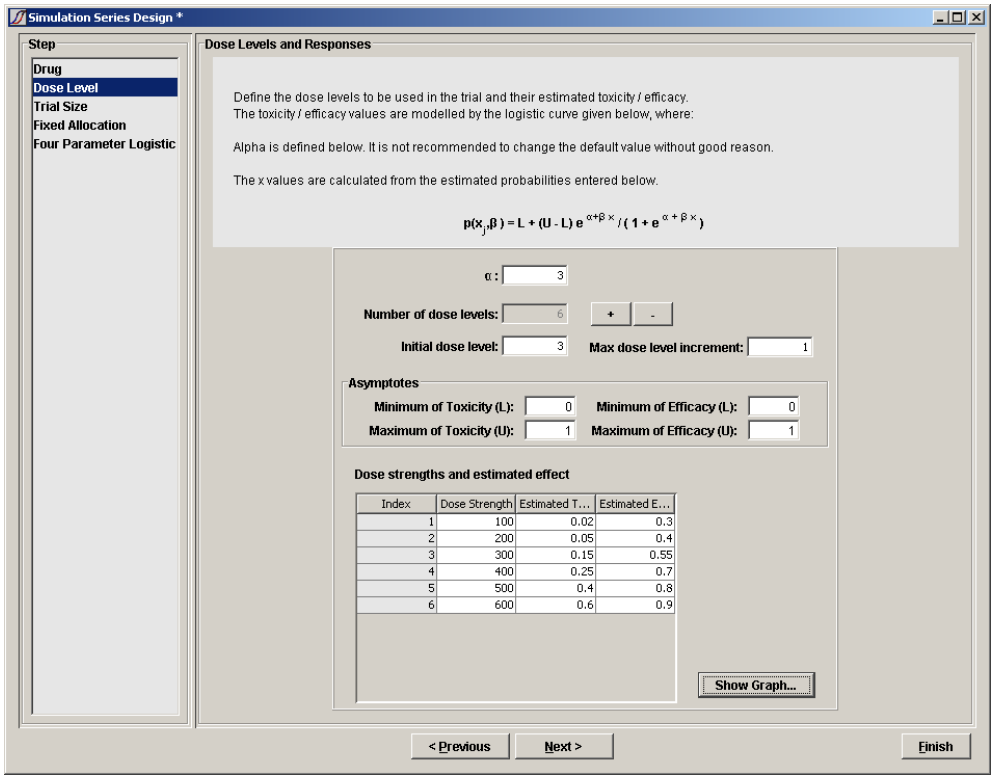


Figure 4-18

Braun appears to use an Alpha of –3 rather than 3, but this has poor properties in the case where the lowest dose is toxic and we prefer to set Alpha to 3.

The toxicity and efficacy priors for the doses are exactly as in [Braun] – except that for efficacy we are using the converse of his disease progression rates.

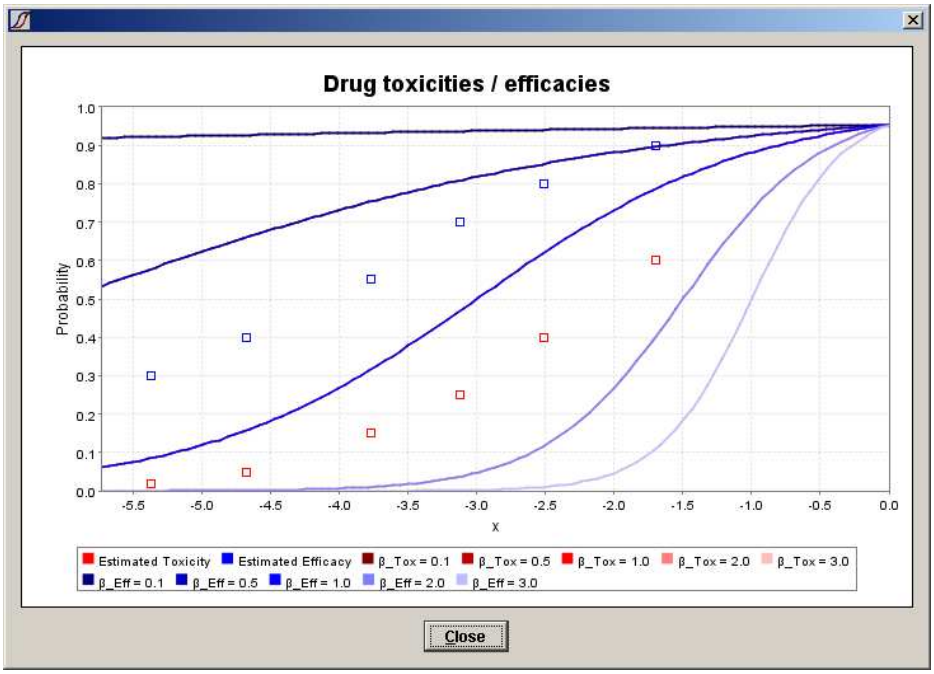


Figure 4-19

For comparison, the family of curves with Alpha set to –3 is shown here:

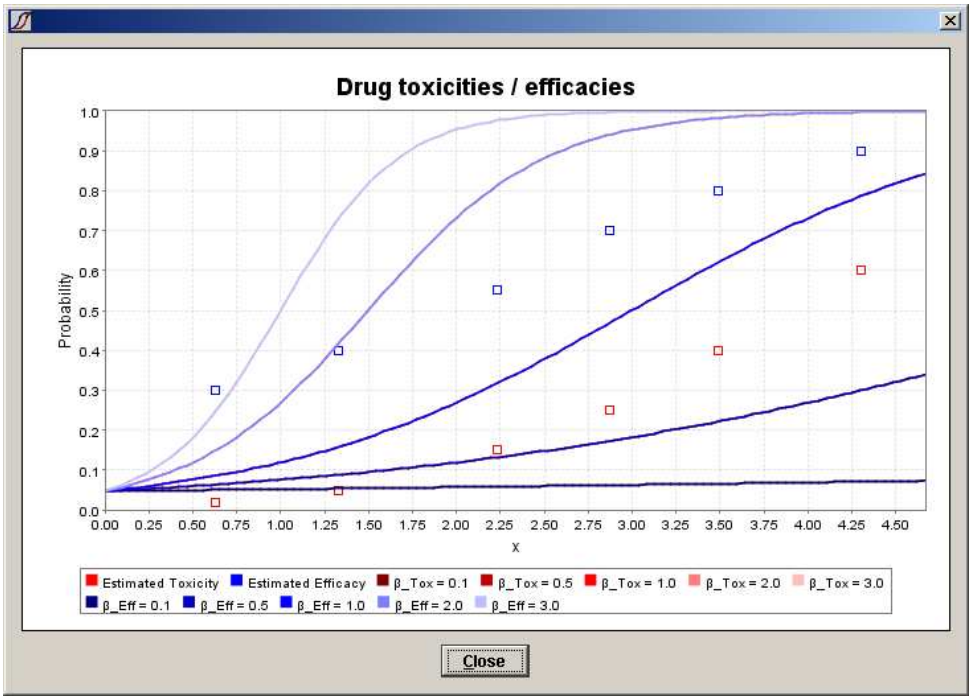


Figure 4-20

Note that using Goodman's uniform prior for Beta (as we do) means that for the highest Beta the curves don't accommodate toxicity above the target at the lowest dose. Presumably Braun avoids this problem by using the exponential prior for beta – which has a larger but non-uniform range.

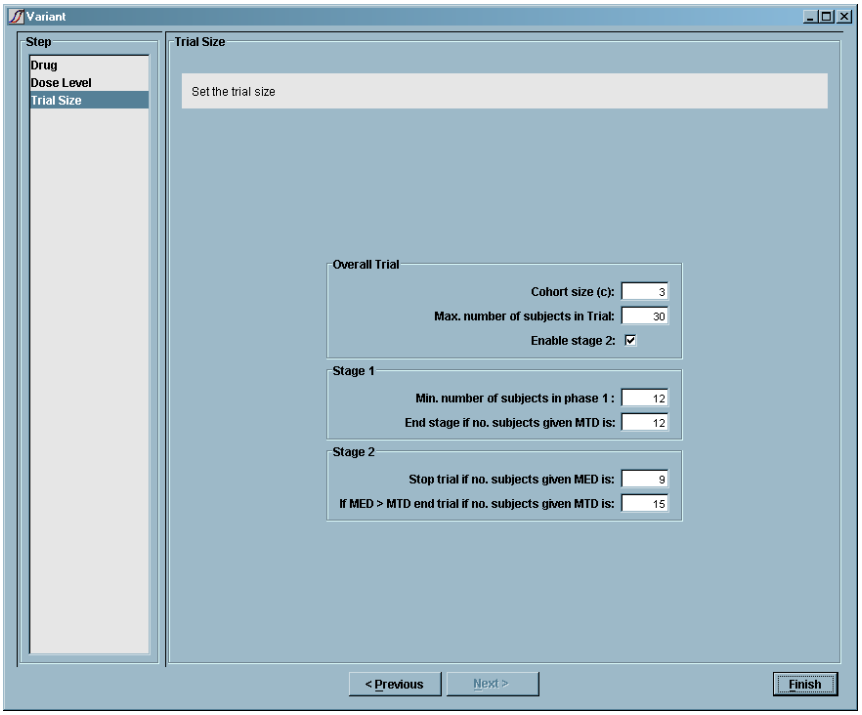


Figure 4-21

Braun uses a study size of 30 subjects and early terminates if no MTD/MED can be found. We allow early termination once there are 12 subjects on the MTD and 9 subjects on the MED (or no available MED <= MTD). We use cohorts of 3 subjects.

4.2.2 Scenarios

These are exactly as in [Braun]:

	Prior Toxicity						
Dose Level	Curve 1	Curve 2	Curve 3	Curve 4	Curve 5	Curve 6	Curve 7
100	0.02	0.02	0.05	0.375	0.02	0.00	0.05
200	0.05	0.05	0.10	0.40	0.05	0.10	0.10
300	0.15	0.10	0.25	0.50	0.15	0.18	0.20
400	0.25	0.25	0.35	0.60	0.25	0.22	0.35
500	0.40	0.60	0.50	0.75	0.40	0.24	0.45
600	0.60	0.85	0.60	0.95	0.60	0.25	0.60

Table 4-5

	Prior Efficacy						
Dose Level	Curve 1	Curve 2	Curve 3	Curve 4	Curve 5	Curve 6	Curve 7
100	0.03	0.10	0.35	0.30	0.05	0.65	0.35
200	0.04	0.25	0.55	0.40	0.15	0.70	0.50
300	0.55	0.35	0.70	0.55	0.30	0.75	0.65

400	0.70	0.70	0.80	0.70	0.45	0.80	0.80
500	0.80	0.90	0.85	0.80	0.55	0.85	0.85
600	0.90	0.95	0.90	0.90	0.65	0.90	0.90

Table 4-6

Simulated Probability of Toxicity/Efficacy	
Curve 1	Curve 2

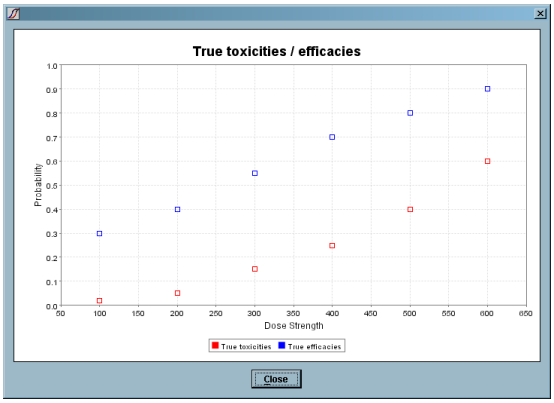
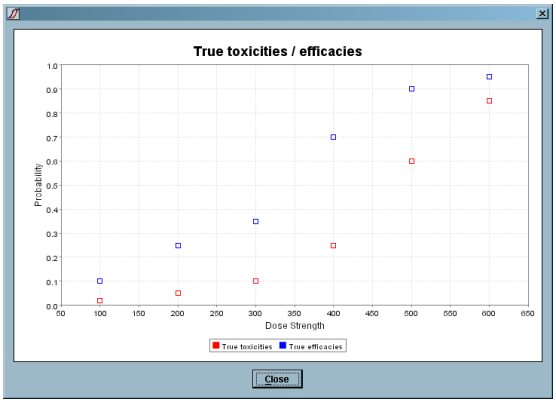

	

Figure 4-22	**Figure 4-23**
Curve 3	Curve 4

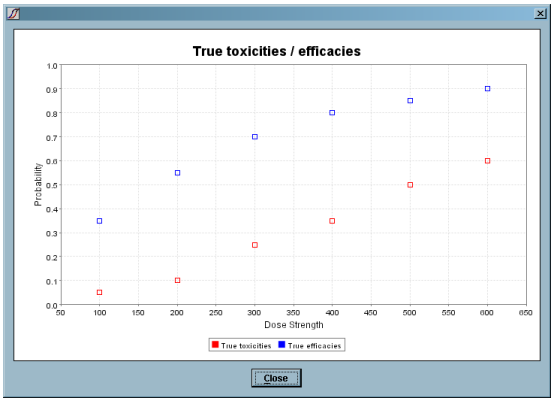
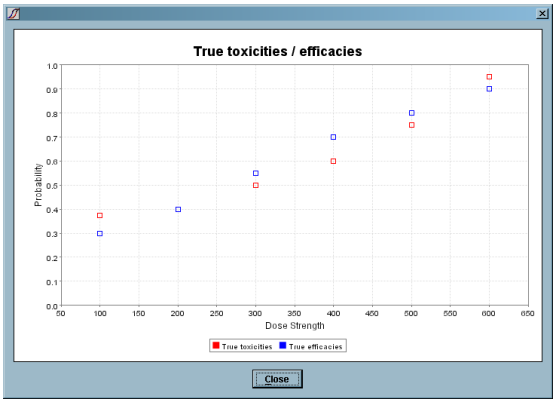

	

Figure 4-24	**Figure 4-25**
Curve 5	Curve 6

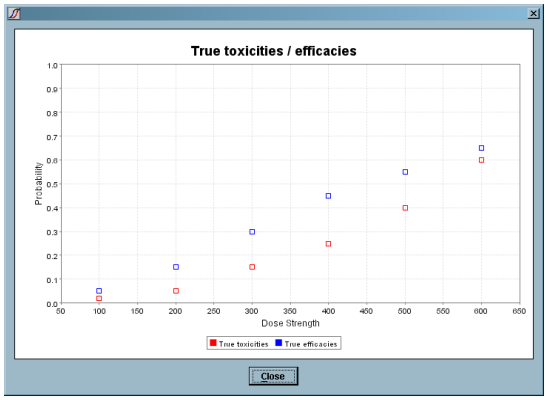


Figure 4-26

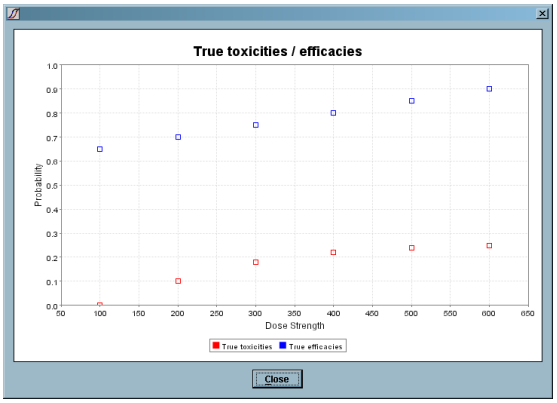


Figure 4-27

Curve 7

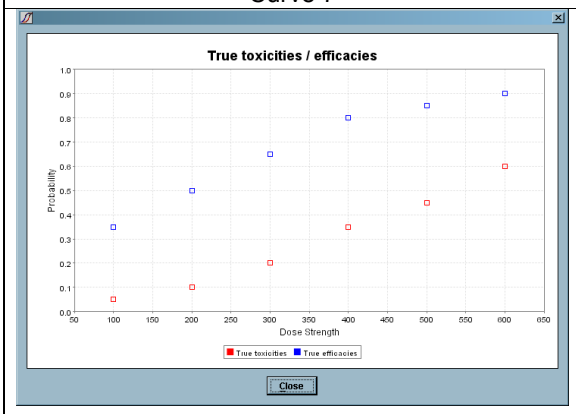


Figure 4-28

Table 4-7

The following is a brief description of the scenarios:

Scenario 1. This scenario conforms exactly to the prior – and there is one dose (400) where the toxicity and efficacy are acceptable.

Scenario 2. This scenario is similar to 1 except the curves are steeper around the acceptable dose.

Scenario 3. This scenario is similar to 1 except the one acceptable dose is at 300.

Scenario 4. The efficacy is as in scenario 1 but there is no acceptable dose for toxicity.

Scenario 5. The toxicity is as in scenario 1 but there is no acceptable dose for efficacy.

Scenario 6. All doses have acceptable toxicity and efficacy. (Under Braun's scheme the lowest dose has unacceptable efficacy, but with our slightly shifted efficacy target it becomes acceptable).

Scenario 7. Similar to scenario 1 but with both curves shifted slightly so that maximum acceptable dose is at 300 and minimum efficacious dose is at 400.

4.2.3 Results

Braun's MTD is actually a combined MTD/MED dose that lies between his target toxicity and efficacy rates, maximising a root mean square metric[1]. Braun ran two lots of simulations, in the first

[1] Braun also presents two more sophisticated metrics, the text is not explicit but I think this, the 'Euclidean Metric', is the one he uses in his simulations.

allocation starts at the expected MTD, in the second allocation starts at the lowest dose (as we do in the bCRM simulations). We report the MTD and MED separately – so the finding the correct MTD results are not directly comparable between the two schemes. The following results are based on the format of the results in [Braun].

		Scenario						
Parameter		**1**	**2**	**3**	**4**	**5**	**6**	**7**
Correct MTD found	Braun 1	46.1%	61.6%	53.8%	n/a	n/a	94.6%	n/a
	Braun 2	23.8%	37.2%	48.0%	n/a	n/a	91.3%	n/a
	bCRM	48.5%	72.6%	52.9%	(75.8%)	53.5%	(5.7%)	49.5%
Correct MED found	bCRM	53.7%	41.1%	52.7%	44.8%	(53.9%)	54.6%	23.1%
Correct MTD +/- one level	Braun 1	82.4%	77.4%	86.8%	n/a	n/a	95.8%	73.6%
	Braun 2	58.0%	52.9%	81.5%	n/a	n/a	97.5%	64.0%
	bCRM	90.9%	98.6%	97.2%	n/a	97.0%	n/a	95.9%
Correct MED +/- one level	bCRM	97.4%	93.2%	89.7%	85.9%	n/a	78.3%	77.8%
No viable MTD found	Braun 1	14.7%	20.3%	11.2%	96.6%	89.6%	4.2%	16.3%
	Braun 2	37.6%	43.7%	15.5%	97.8%	97.0%	2.5%	22.4%
No viable [MED,MTD]	bCRM	30.1%	56.2%	25.6%	99.1%	92.5%	1.7%	27.6%
Average Toxicities per trial	Braun 1	7.0	7.9	7.3	6.4	4.9	5.0	7.0
	Braun 2	6.4	7.6	6.8	5.1	3.3	4.3	6.5
	bCRM	5.6	5.2	5.5	5.8	5.0	4.2	5.2
Average Efficacies per trial	Braun 1	18.2	17.6	19.7	6.7	6.8	22.0	18.9
	Braun 2	16.4	14.3	18.1	4.8	4.2	22.0	17.1
	bCRM	16.4	13.5	16.6	5.0	8.9	21.4	16.2
Average subjects per trial	Braun 1	27.6	27.2	28.3	12.6	14.2	29.3	27.6
	Braun 2	26.4	24.6	27.3	9.8	10.6	29.5	26.4
	bCRM	27.9	27.2	26.4	14.9	26.8	28.4	26.3

Table 4-8

In general, the bCRM finds the exact MTD roughly 50% of the time, It does better in scenario 2 where the curve at the MTD point is steep and scenario 4 where the MTD doesn't exist because all doses are too toxic, getting it exactly right ~75% of the time. In scenario 6 where the toxicity curve is very shallow and all are acceptable, it finds an MTD typically around dose 4 (when there are 4 doses with toxicity of 20%, the chances that one of them will appear to have toxicity of over 30% is very high).

The MTD is found to within one dose level ~95% of the time (with the exception of scenarios 4 and 6 where it is not defined) and the MED ~90% of the time. The MED is found less accurately than the MTD because fewer subjects are dedicated to finding the MED and if the MTD is low it may prevent the actually MED being assigned to subjects.

Because once it has found an MTD it will not go above it to look for an MED, bCRM tends to see slightly less toxicity and less efficacy than the Braun algorithm (it tends to allocate lower doses).

The bCRM tends to fail to find a viable [MED, MTD] interval more often than the Braun method. But Braun's viability criteria are less stringent and an MTD is only non-viable if there is a 95% confidence that both the overall toxicity and efficacy will fail to meet the targets.

This test data is included on the bCRM CD as the simulation series 'Thomas-Braun-Example'. The results presented here are from the design variant 'target-30-60-late'. Two other variants are included:

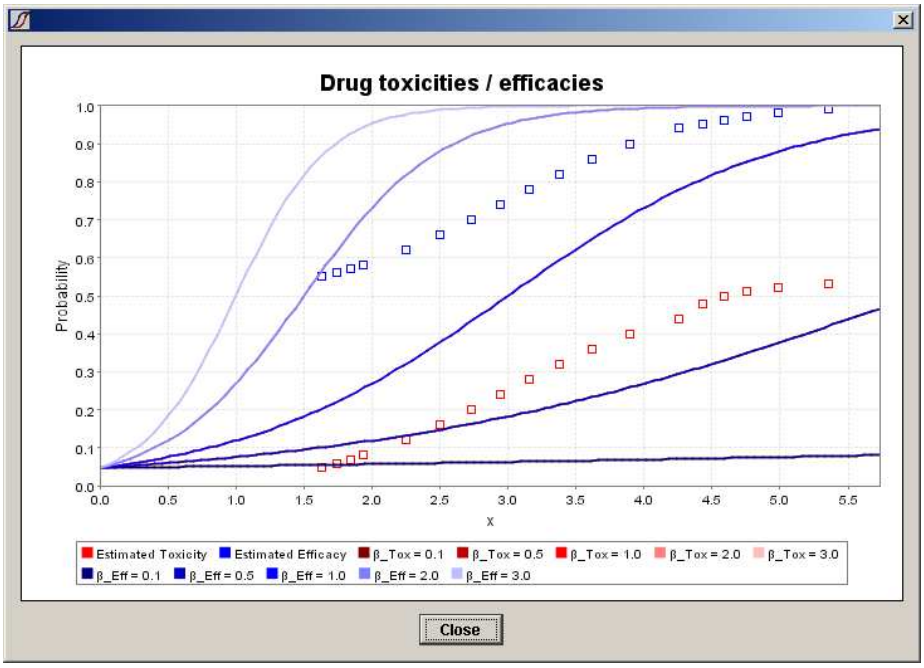
target-30-60-early	as the presented data but with fewer subjects required on the MTD and MED before termination is allowed – uses fewer subjects, but is less accurate.
target-30-65-early	as above but has a target efficacy of 65%, this is not enough of a change of target (from 70%) to offset the algorithm looking for the minimum dose *above* target rather than the *nearest* dose to the target.

4.3 Step function toxicity and efficacy examples from [Braun]

Braun includes some examples from a paper by Gooley et al. However the simulated data in this example is a poor candidate for the parametric curve of this method:

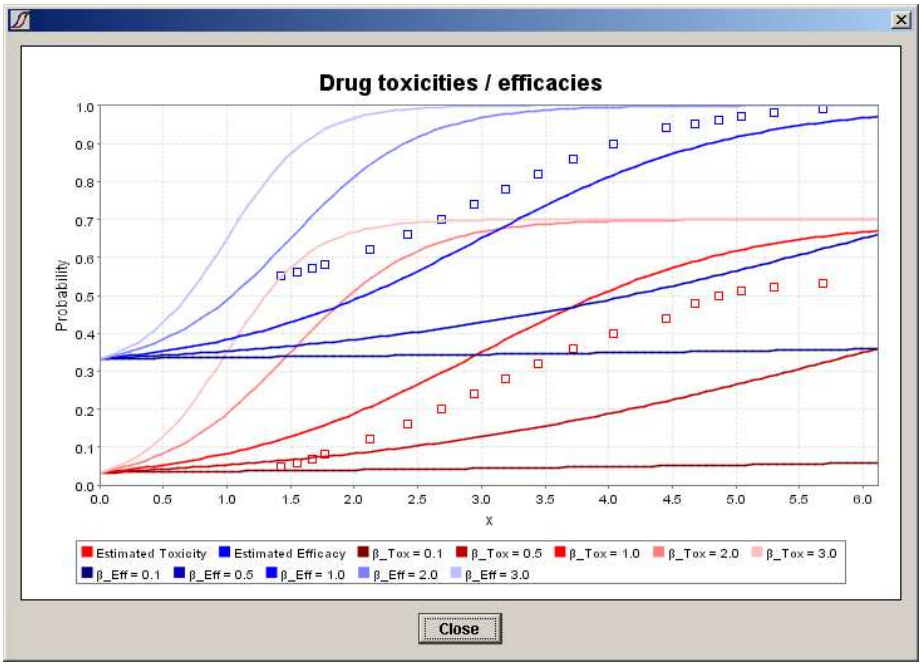
1. The scenarios have steep 'step' like changes from low toxicity to high toxicity and low efficacy to high efficacy that lie well outside the families of curves achievable with the logistic.
2. The upper value of the toxicity curve is 0.5, the logistic models a curve that goes from 0 to 1.
3. Many doses are used, with the result that they lie close together, and the sharp change in toxicity that occurs between doses in the simulated scenarios cannot be represented. The logistic model therefore introduces a smoothing that destroys information rather than adds to it.

[Our bCRM performs similarly to Braun's using the same setup, but performs better (in terms of finding the valid interval) if we start at dose 3 instead of dose 14.]



The graph above clearly shows how the beta=1 line lies between the estimated toxicities and efficacies but is close to neither and has the wrong shape. Setting an upper limit on toxicity of 0.7 and a lower limit efficacy of 0.3 makes the parametric curve a much better fit to Gooley's curves.

The graph below shows the beta=1 toxicity and efficacy lines lying much closer to their respective points and having the correct shape.



Results are also much improved. The scatter in MTD is reduced by 50% and the scatter in MED by 30% compared to asymptotes of 0 and 1.

This test data is included on the bCRM CD as the simulation series ‘Thomas-Braun-Gooley-Example’. Result for the following variants are included:

target-15-95-start-dose-3	This uses the targets exactly as in the example, but start from dose 3. The algorithm performs poorly with these parameters. The targets need to be offset to allow for the algorithm looking for the minimum dose *above* target rather than the *nearest* dose to the target.
target-20-90-start-dose-3	This uses targets adjusted to allow for the difference in algorithm and a low starting dose.
target-20-90-start-dose-14	This uses targets adjusted to allow for the difference in algorithm and the same, high, starting dose as in [Braun].
new	Version with changed asymptotes.