# ToxFinder – Case Study

## Tessella Project Number 3760

# ToxFinder Case Study: Reproducing the Paper


**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

### Contributors

| Person | Role | Company | Contribution |
|---|---|---|---|
| Robert Nelson | Author | Tessella | |
| | Reviewer | Tessella | |
| | | | |

### Document Approval

| Person | Role | Company |
|---|---|---|
| | | |
| | | |

### Distribution

| Person | Role | Company |
|---|---|---|
| Peter Thall, Peter Muller | | MD Anderson |
| Notes Projects Database 3760 | | Tessella |

### Revision History

| Issue | Author | Date | Description |
|---|---|---|---|
| V1.R1.M0 | Robert Nelson | 13-May-2004 | 1st version written |
| V1.R2.M0 | Nick Battam | 18-Jan-2005 | Updated to reflect revised user interface |
| V1.R2.M1 | Tom Parke | 19-Jan-2005 | Re-organised section 5.1 and removed section 6. |

### References

| Ref. | Document | Date | Details and Version |
|---|---|---|---|
| [TMML] | Dose-Finding with Two Agents in Phase I Oncology Trials | Sept-2003 | P.F.Thall, R.E.Millikan, P.Mueller and S-J.Lee. (2003). Biometrics, **59**, 487 |
| [SUG-S] | "ToxFinder – Defining and Running Simulations" – System User Guide | 03-Dec-2004 | NPD/3670/SUG "ToxFinder – Defining and Running Simulations" V1.R4.M0 |

# Table of Contents



# 1 Introduction

## 1.1 Purpose of this document

This document describes, through an example, how to use the ToxFinder software v1.0.0. The example is based on that described in [TMML]. It is intended for all end users of the system.

## 1.2 Scope of this document

This document is not a detailed User Guide; there are separate System User Guides document that fulfils this role.

This document describes setting up and simulating a combinatorial phase I oncology trial and examines the effects of changes to the prior P(Toxicity) surface and allocation methods.

## 1.3 Context of this Issue

This is the second released version of this document, updated to reflect changes made to the ToxFinder user interface.

## 1.4 Definition of Terms

ToxFinder — is a pair of computer programs that implement the TMML dose-finding algorithm, simulate and run clinical trials using the TMML design and provide charts and graphs of the results.

A Simulation Series — is comprised of a 'Design' for a trial using TMML, a set of one or more 'Scenarios' – possible probabilities of toxicities – 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 set of parameters that define a trial design to be carried out using the TMML method. This consists of: the two dose ranges to be used, lambda - the relative cancer killing potential of the two drugs, the prior distribution of the parameters that characterise the probability of toxicity as a function of the dose pair of the two agents, the cohort size, the sample sizes of the two stages of the trial, the fixed line L1 used in the first stage of the trial and the first set of doses to use on L1.

A Scenario — is a set of fixed values of the probability of toxicity as a function of the two agents' doses, $\mathbf{x} = (x_1, x_2)$. We denote this probability by $\pi(\underline{\mathbf{x}})$. As $\mathbf{x}$ varies over the two-dimensional domain of the dose pairs $\pi(\mathbf{x})$ forms a surface that may be illustrated graphically. ToxFinder does this, providing a graphical representation of each scenario specified by the user.

A Variant — is a small modification to a Design – e.g. different prior, cohort size or study sample size.

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 consists of 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.tmml' 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 file format is readily imported into many other programs such as Excel, Access, SAS and Oracle). ToxFinder 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 ToxFinder. ToxFinder creates data for subjects and their responses by randomly selecting from the relevant probability distributions of the scenario defined in the ToxFinder 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 usually 1,000-10,000 times over a range of scenarios. |
| Trial | A trial is a real clinical trial, where the user enters subjects' data and responses. The ToxFinder Algorithm is run to determine the doses to allocate and to analyse the results. This facility allows ToxFinder's adaptive allocation to be implemented to conduct small (single centre) trials. |
| Cohort | Subjects are treated in small groups called cohorts. The subjects in a cohort all receive the same doses of the two drugs. Each subject either experiences toxicity or does not experience toxicity as a result. |

## 1.5 References

[TMML] "Dose-Finding with Two Agents in Phase I Oncology Trials." P.F.Thall, R.E.Millikan, P.Mueller and S-J.Lee. (2003). Biometrics, **59**, 487

# 2 The System User Guide

The System User Guide [SUG-S] contains an introduction to the system, instructions on how to install the system and detailed descriptions of each screen.

Please refer to the SUG for any of these details. This document just contains some examples of use.

# 3 Summary of Findings

The case study shows that adaptive dose finding can be a practical proposition.

The simulations show the system able to determine the shape of the dose/toxicity surface.

The simulations show the system to be insensitive to variations in the prior over a wide range.

# 4 Setting up the Study

This simulated study was based on a proposed study described in [TMML]. In some cases values have been estimated from graphs.

The simulated study looks at the combined effect of using a combination of Gemcitabine and Cyclophosphamide.

The protocol details to be modelled in this simulation series are:

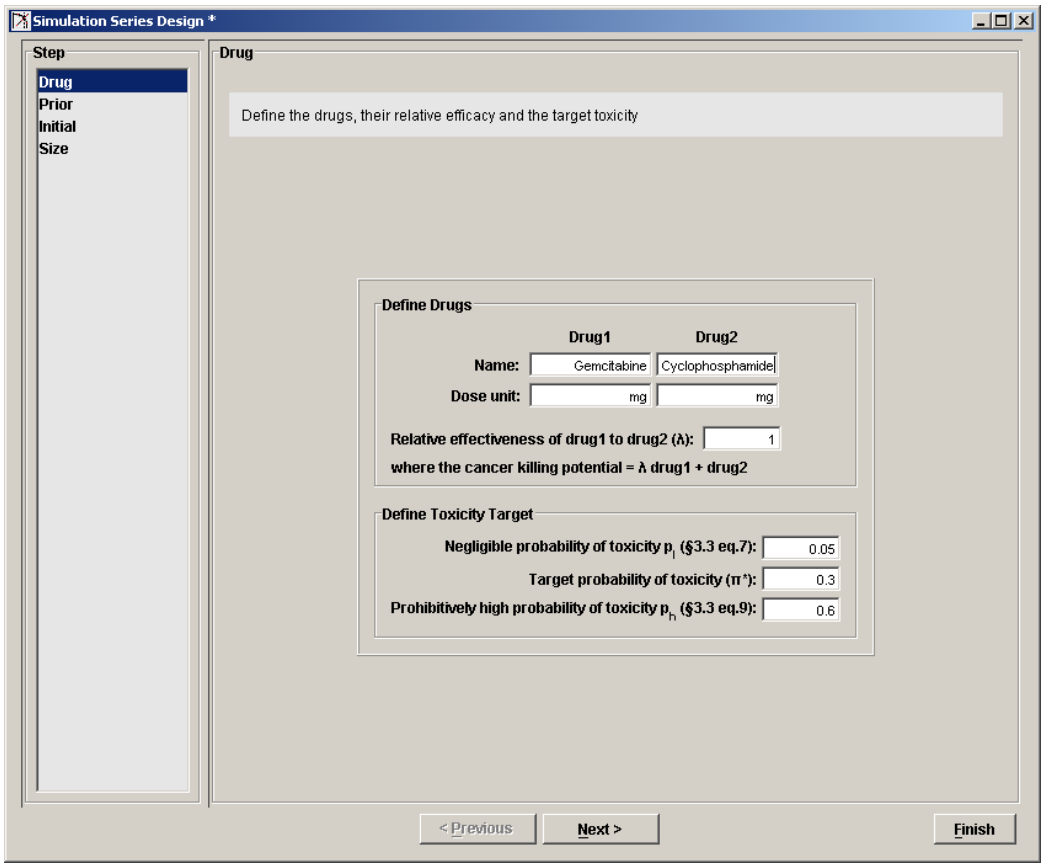
| | |
|---|---|
| Drug | 2 drugs, Gemcitabine and Cyclophosphamide with maximum tolerable doses of 1200mg and 600mg (defined as 30% of subjects suffering toxicity). The objective is to find a combination of the two drugs, which has similar toxicity but is more effective than either used alone. It is assumed that both are equally effective. |
| Prior Response | The priors are specified by the probabilities given in [TMML] and the doses listed in [TMML] table 1. |
| Initial | The first stage of the trial involves escalating the doses of the two drugs together to find a maximum tolerable dose for the two together. Since the maximum tolerable doses of the two alone have a ratio of 2:1, we will use the two drugs in this ratio during stage 1. The starting dose combination, which we believe will not be toxic, is chosen as 225mg and 112.5mg. |
| Study Size | The study is sized as described in [TMML]. |

A number of scenarios are specified with a range of simulated P(Toxicity) surfaces, listed below.

The sub-sections below show how the above parameters appear in the ToxFinder parameter editing screens.

## 4.1 Series: Drug

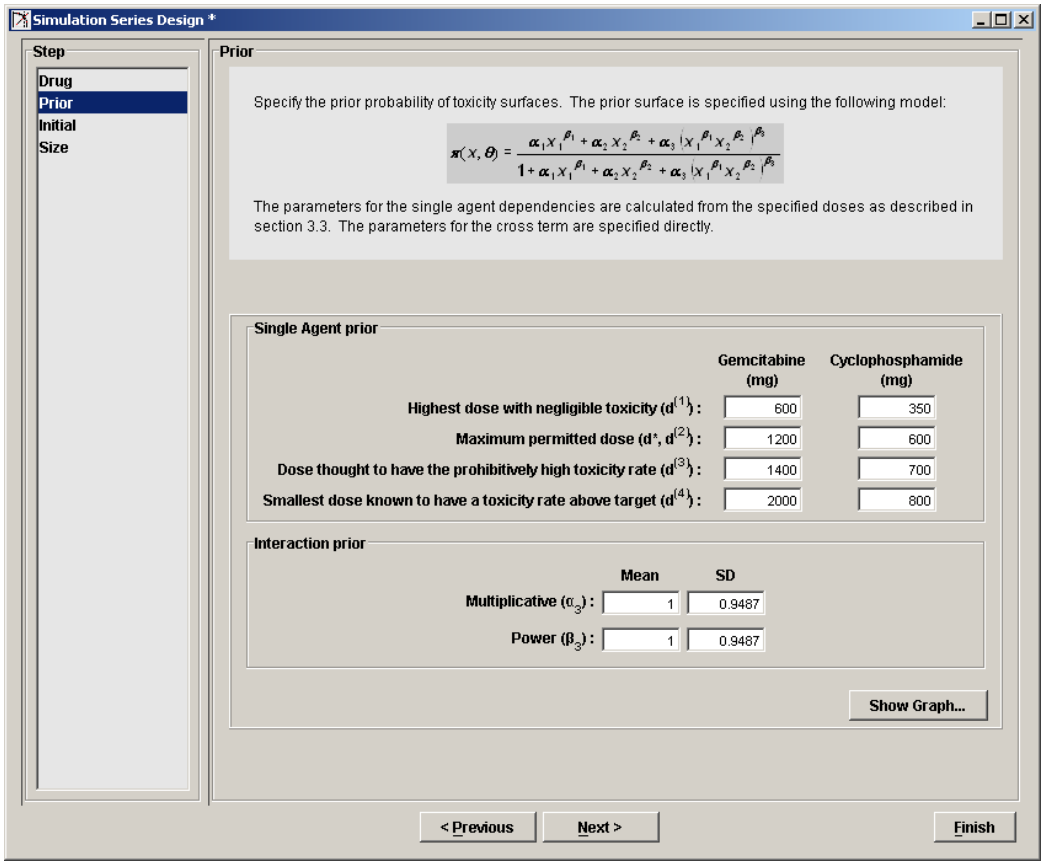
The target toxicity is 30%, which corresponds to doses of 1200mg and 600mg of Gemcitabine and Cyclophosphamide alone. The two drugs are assumed to have equal cancer killing effect at these doses.



## 4.2 Series: Prior

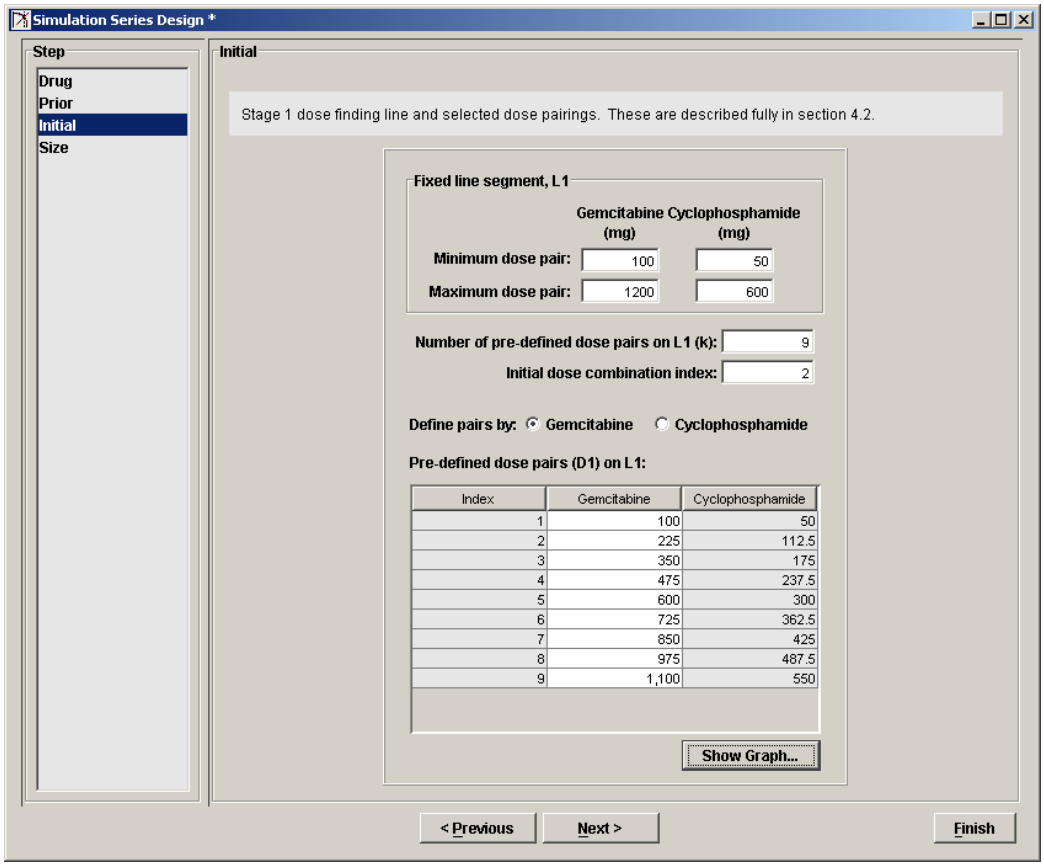
The priors are taken from [TMML] table 1. The only difference is that power term in the interaction prior ($\beta_3$) has been changed from 0.05 with a variance of 3 to 1 with a variance of 0.9. The reason is that original prior encouraged solutions in which $\beta_3$ was very close to zero, resulting in a surface with a plateau. The new prior encompasses the expected range for $\beta_3$ while excluding zero. This will be discussed in more detail below.



## 4.3 Series: Initial

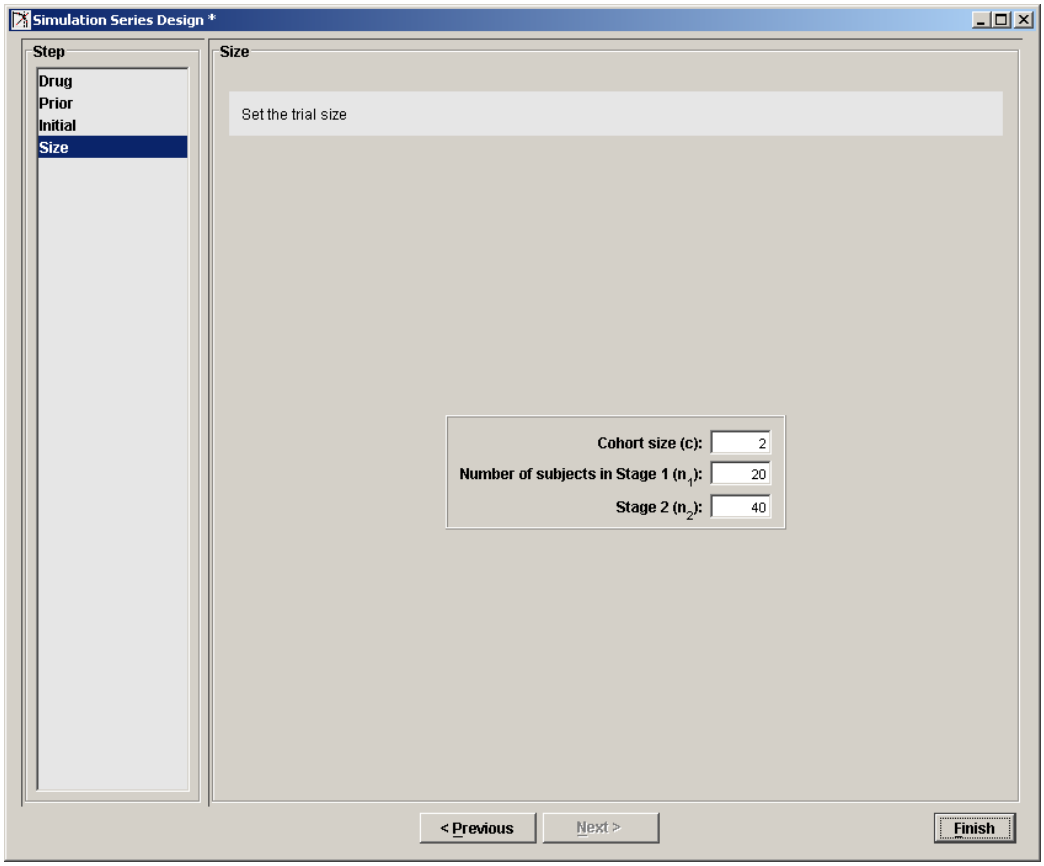
During stage 1, dose allocation is restricted to specified dose combinations which here lie along a diagonal line on a dose/dose graph. Nine specific combinations are defined. Allocation will start at the second combination (225, 112.5). During stage 1 the algorithm is not allowed to skip untried combinations in this list and after the first toxicity is encountered, it is also required to step through additional combinations placed halfway between each pair in the initial list.



## 4.4 Series: Size

Here we set the cohort size and the numbers of subjects in each of the two stages. Again values are taken from [TMML]

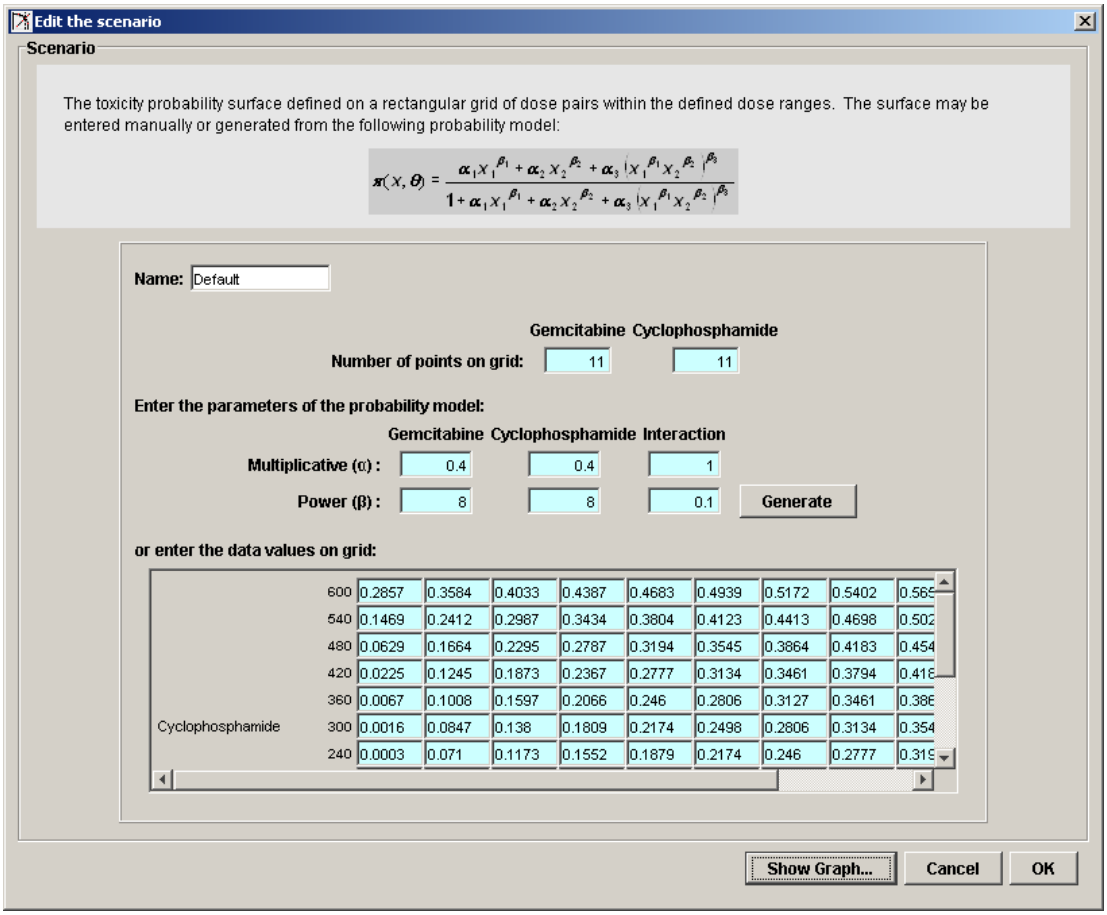


## 4.5 Scenario Definitions

Here we define the true toxicity surface, which the system will attempt to discover. The grid specifies the probability of toxicity for different dose combinations. The size of the grid can be changed and if the values are going to be entered manually, the size can be set to something smaller than 11x11 to reduce the amount of typing. Alternatively values can be entered for the multiplicative and power terms and the generate button can then be used to populate the surface automatically (these values are the α's and β's in [TMML] equ. (3)).

Because the algorithm assumes that the probability of toxicity increases monotonically with increasing dose of either drug, the true surface should do the same. If it does not then results are likely to be erratic and inconsistent.

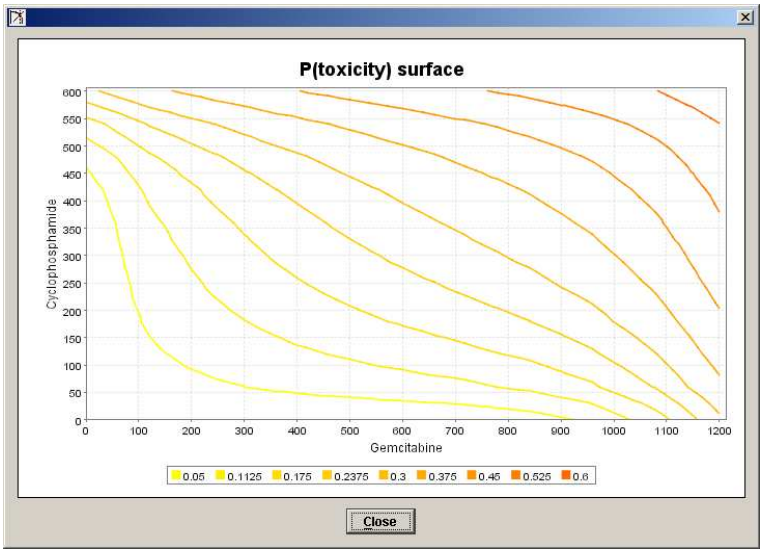



To investigate how well the system performs, we define four true surfaces:

### 4.5.1 Standard

The surface is generated with the alpha's and beta's of [TMML] equation (3) set to:

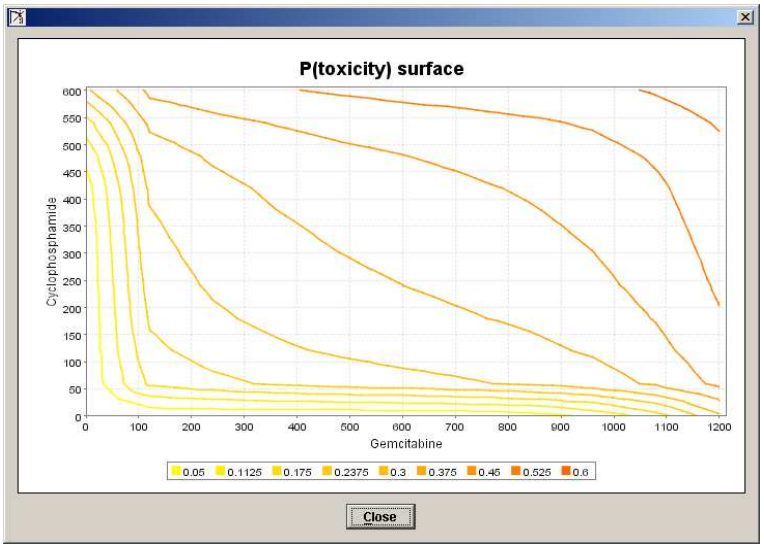
| | Gem | CTX | Interaction |
|---|---|---|---|
| alpha | 0.4 | 0.4 | 1.0 |
| beta | 8.0 | 8.0 | 0.1 |




### 4.5.2 Concave

The surface is generated with the alpha's and beta's of [TMML] equation (3) set to:

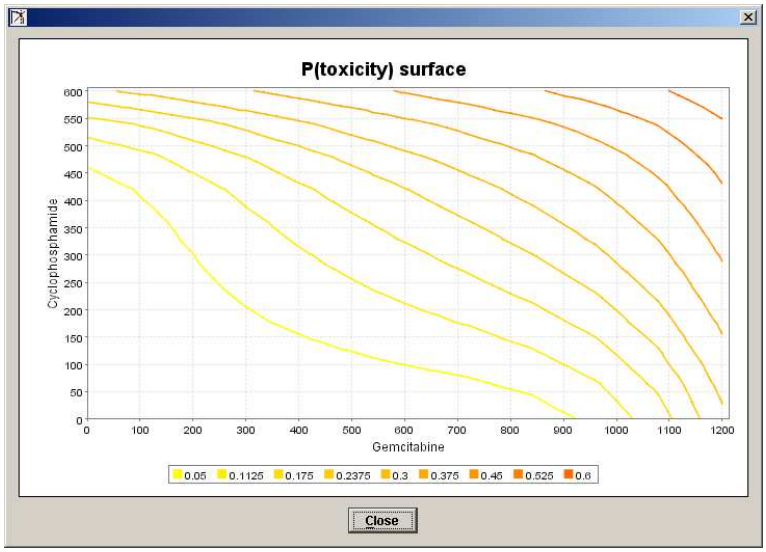
| | Gem | CTX | Interaction |
|---|---|---|---|
| alpha | 0.4 | 0.4 | 1.0 |
| beta | 8.0 | 8.0 | **0.04** |



### 4.5.3 Convex

The surface is generated with the alpha's and beta's of [TMML] equation (3) set to:

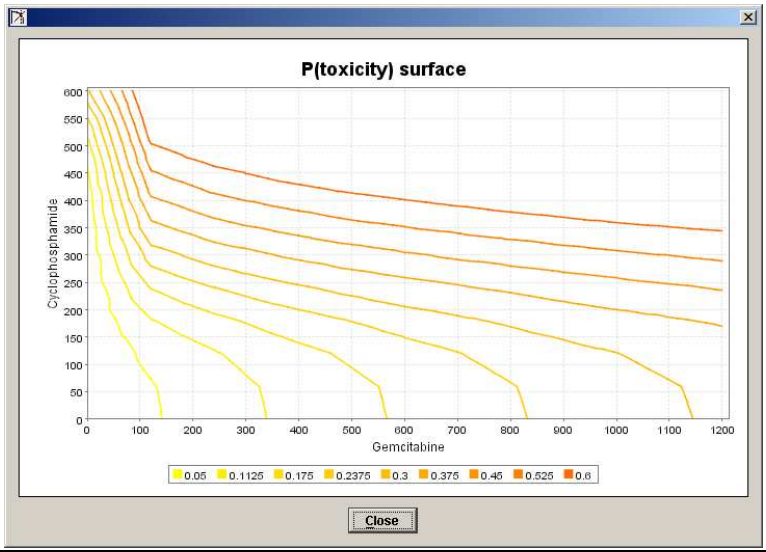
| | Gem | CTX | Interaction |
|---|---|---|---|
| alpha | 0.4 | 0.4 | 1.0 |
| beta | 8.0 | 8.0 | **0.15** |



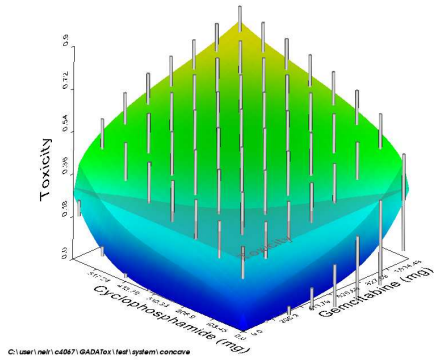
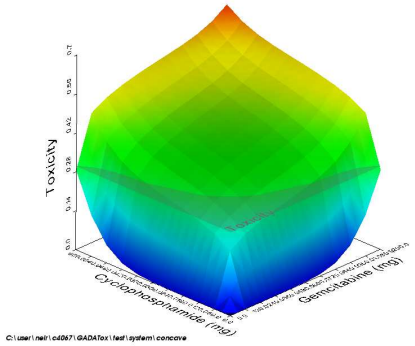
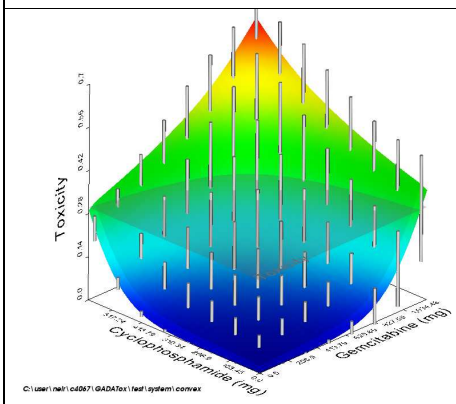
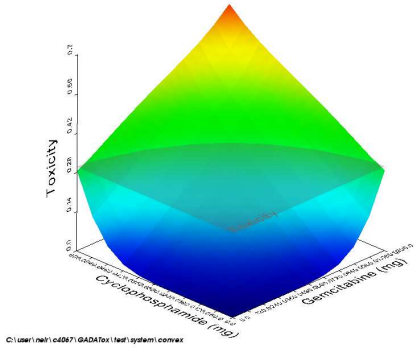
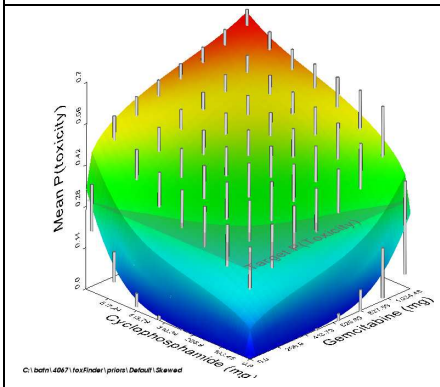
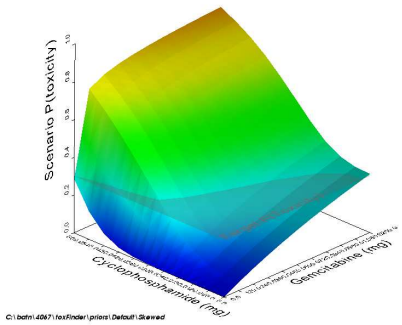

### 4.5.4 Skewed

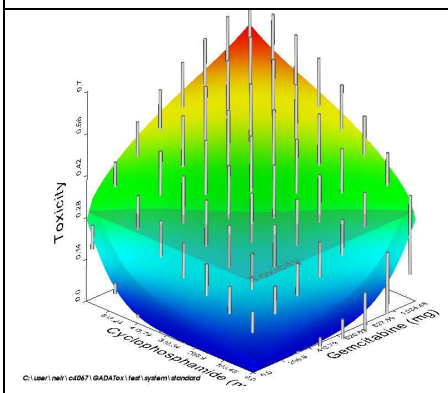
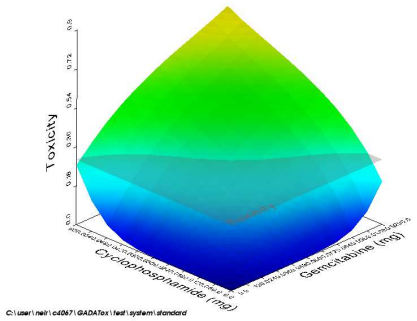
The surface is generated with the alpha's and beta's of [TMML] equation (3) set to:

| | Gem | CTX | Interaction |
|---|---|---|---|
| alpha | 0.4 | 0.45 | 7.5 |
| beta | 8.0 | 1.0 | 0.375 |



To illustrate the systems behaviour ten simulations were performed in each case (this is not nearly enough simulations to give data on which claims could be based). The following graphs show the average surfaces found (left) compared to the true surfaces (right). The horizontal plane is at 30% toxicity.

| Found surface | True surface |
|---|---|
|  |  |
| Concave | |
|  |  |
| Convex | |
|  |  |
| Skewed | |

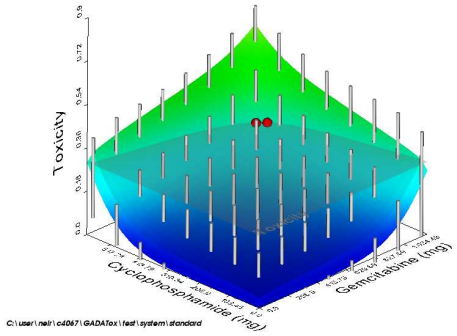
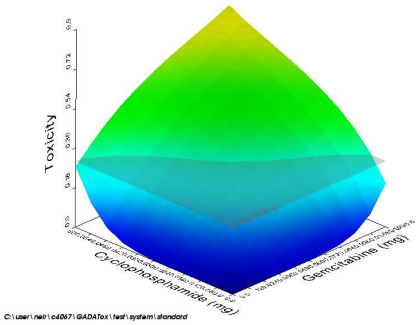
| Found surface | True surface |
|---|---|
|  |  |
| Standard | |

In all four cases the shapes are similar. The biggest difference is for skewed which at first glance has major differences. However these differences are smallest at the 30% level where most interest lies. This is not surprising since the adaptive allocation has assigned most subjects to this region.

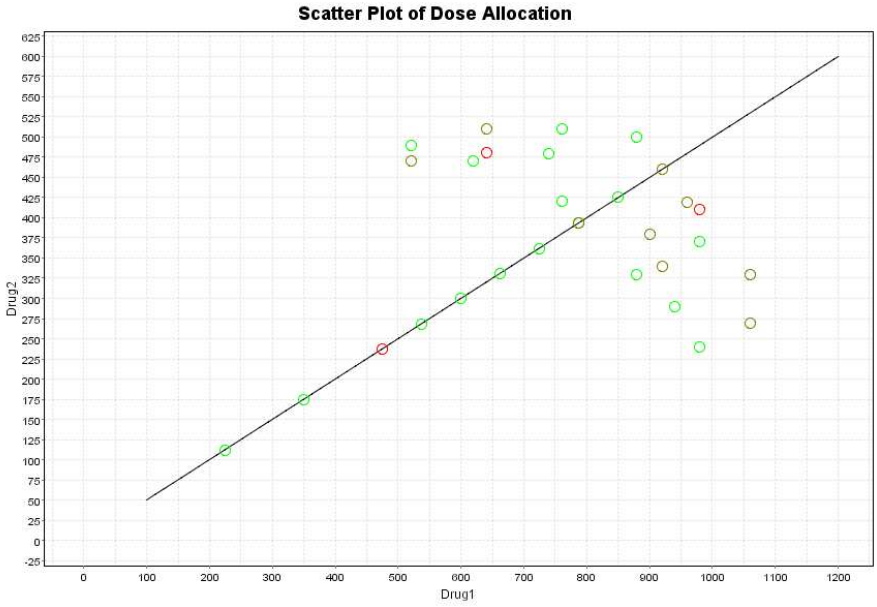
## 4.6 Variability between simulations

Showing averaged results does not necessarily give a true picture of how well the system is working. A real clinical trial would only take place once, so a good the average is no help if the real trial is an outlier. What we also need to assess is the spread of possible results.

If we look at the individual simulations for the standard surface, there are a range of surfaces. The most extreme is simulation 9 where toxicity has been underestimated:

| Simulation 9 | True surface |
|---|---|
|  |  |

The recommended dose on the diagonal is 920, 420mg whereas the true value is 560, 280mg. 920, 460mg corresponds to 50% toxicity rather than the desired 30%. How did the answer come out so wrong? If we look at the scatter plot of dose allocation we see:




Most subjects have been allocated to high doses but the preponderance of green circles shows they have not suffered the expected number of toxicities. Given this data, the choice of 920, 460mg is entirely reasonable. The problem is that the subjects have been lucky and few have suffered toxicity. If we had reason to believe that this was an unlikely outcome, we could adjust the prior so that even stronger evidence would be required to reach this result. In general however we do not know what interaction to expect so the only way to reduce the spread of outliers is to increase the number of subjects.

# 5 Sensitivity to Prior

The results above were obtained using prior values taken from [TMML] with the exception of $\beta_3$. We now check how important the priors are.

## 5.1 Changing $\beta_3$.

In [TMML] it is stated that the prior for $\beta_3$ should be vague and a prior is suggested with a mean of 0.05 and a variance of 3. We have come to prefer and recommend a prior mean of 1 and a variance of 0.9.

This change causes a change to the shape of the prior surface. The mean of the original prior corresponded to a moderately strong interaction between the two drugs, whereas the mean of the new prior corresponds to almost no interaction between them. In other words the new prior is less conservative than the old prior and we will need to check that this change has not made the program unsafe.

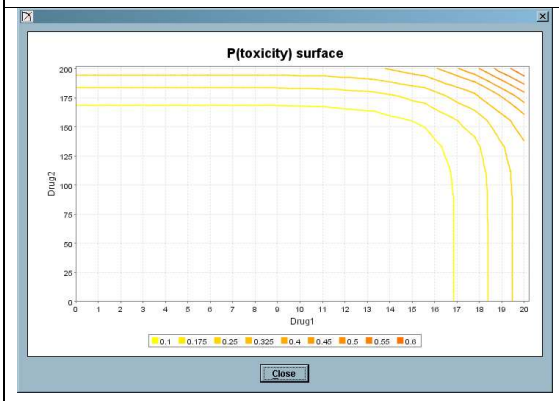
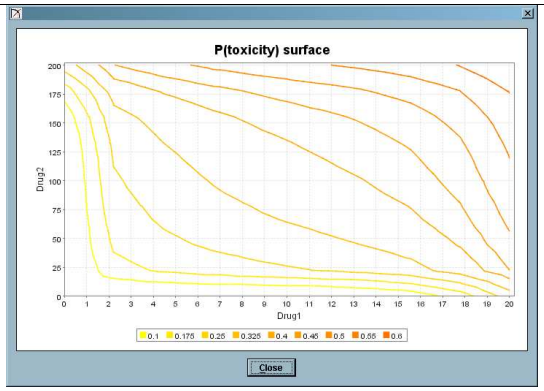
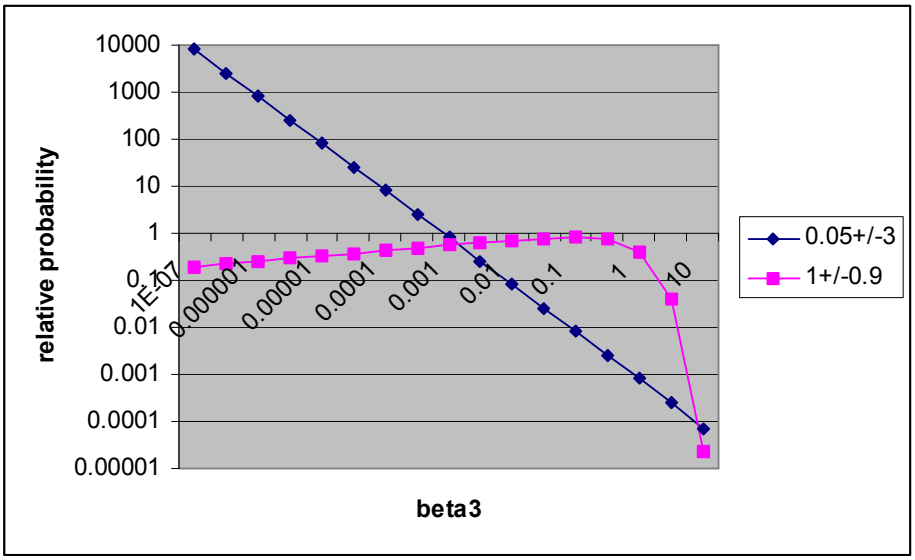
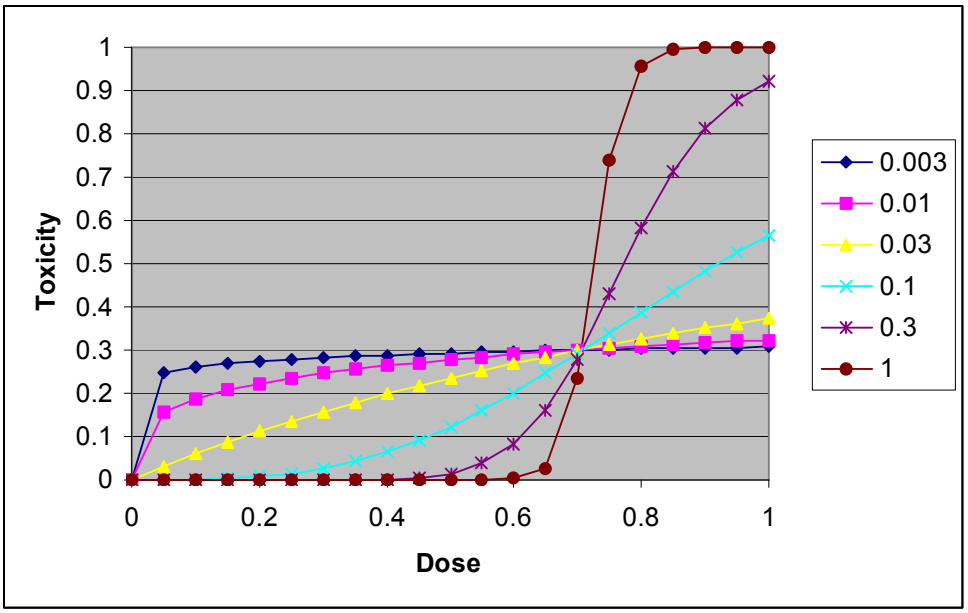
| $\beta_3 = 1$ | $\beta_3 = 0.05$ |
|---|---|
|   |   |
| **Figure 5-1** | **Figure 5-2** |
| $\alpha_1,\alpha_2 = 0.4$, $\alpha_3 = 1$, $\beta_1,\beta_2 = 0.8$ | |

**Table 5-1**

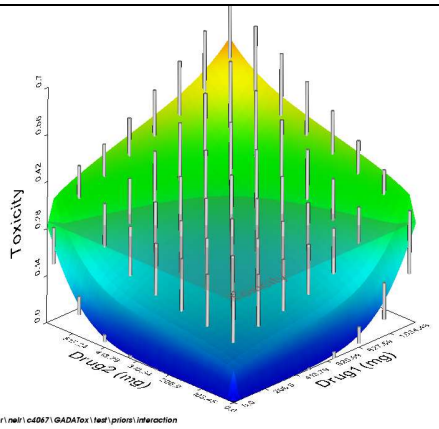
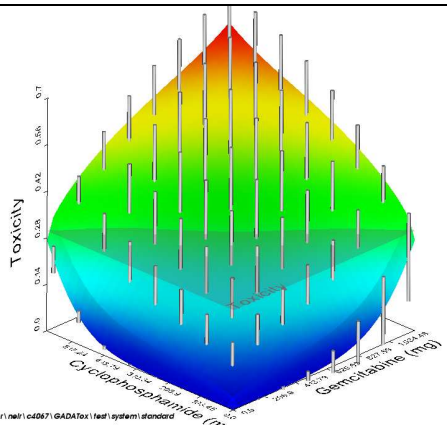
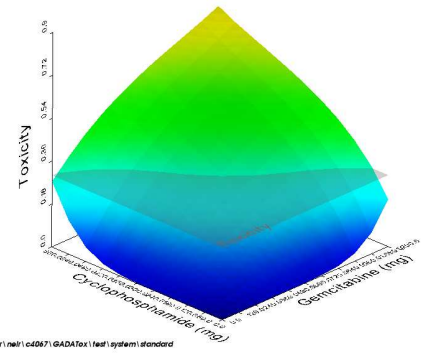
The following graph shows the probability densities for $\beta_3$ under these two priors. As can be seen, the probability density for $\beta_3$ using the prior given by [TMML] tends to infinity as $\beta_3$ tends to zero. Although the cumulative distribution remains finite, the shape of the probability density leads to practical problems, which motivated the switch to a prior with a more evenly distributed probability density.



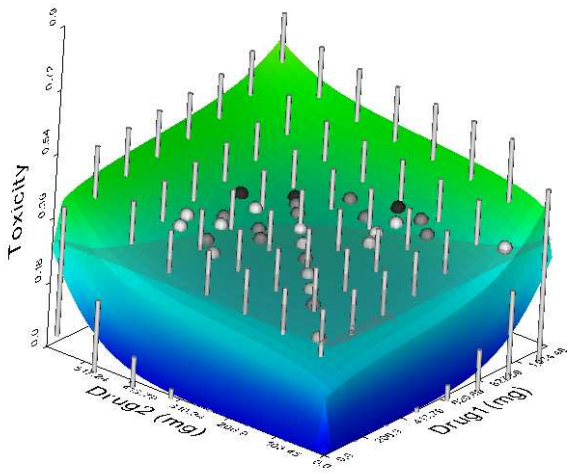

The following graph shows how the shape of the interaction term varies as a function of $\beta_3$. The graph assumes that for simplicity $\alpha_1$ and $\alpha_2$ are zero while $\beta_1$ and $\beta_2$ are ~8 as in [TMML]. The graph is plotted along the diagonal and in each case $\alpha_3$ has been adjusted so that 30% toxicity occurs at a dose of 0.7 of both drugs. As can be seen $\beta_3$=1 gives an unrealistically abrupt onset of toxicity while at the other extreme $\beta_3$=0.003 gives an almost constant interaction. Since few if any subjects receive doses of less than 0.1, for small $\beta_3$, the interaction term is essentially constant over all subjects controlled by the value of $\alpha_3$. Because of the shallowness of the slope over the effective part of the range, the location of the 30% toxicity boundary becomes very sensitive to the value of $\alpha_3$.



The mean of the original prior at 0.05 corresponds to a sensible interaction, but its probability density is highest for small values, which effectively give no interaction. Thus the original prior is not as conservative as it appears to be. The standard case above (4.5.1) was rerun with the prior taken from [TMML]. As can be seen below, this results in fitted surfaces with less interaction between the two drugs than either the true surface or the simulations run with the new prior..

| | |
|---|---|
|  | TMML:<br>Mean = 0.05<br>Variance = 3 |
|  | Alternative:<br>Mean = 1<br>Variance = 0.9 |
|  | True surface |

For example simulation 2 when using the TMML prior:




In this case $\beta_3$ has become close to zero. This results in the interaction between the two drugs occurring for very low doses and giving the steep rise in toxicity seen along the edges of the surface. At higher doses the interaction term has reached its maximum and results in a plateau.

## 5.2 Strength of Prior Surface

The priors of α and β are modelled with Gamma distributions. The expected values of these parameters are found analytically, however the variances must be calculated iteratively by drawing a large number of samples from surfaces generated with candidate distributions of α and β.

The toxicity rates at the three input doses:

'Highest dose with negligible toxicity $d^{(1)}$,

'Dose **thought** to have the prohibitively high toxicity rate $d^{(3)}$,

and 'Smallest dose **known** to have a toxicity rate above target $d^{(4)}$,

are compared with the specified probabilities ('Negligible probability of toxicity, $p_l$', 'Prohibitively high probability of toxicity, $p_h$' and 'Target toxicity, $\pi^*$') and the candidate distributions updated.

Empirical relationships that characterize the effects of changes to the input doses, $d^{(1)}$, $d^{(3)}$ and $d^{(4)}$ on the variance of α and β have been derived, where $d^{(1)} < d^* < d^{(3)}, d^{(4)}$.

| Increase | var(α) | var(β) |
|---|---|---|
| $d^{(1)}$ | + | -- |
| $d^{(3)}$ | - | --- |
| $d^{(4)}$ | ++ | - |

In this table: '**-**' represents a reduction in the derived variance and '**+**' represents an increase in the variance. Repeated symbols represent larger effects.

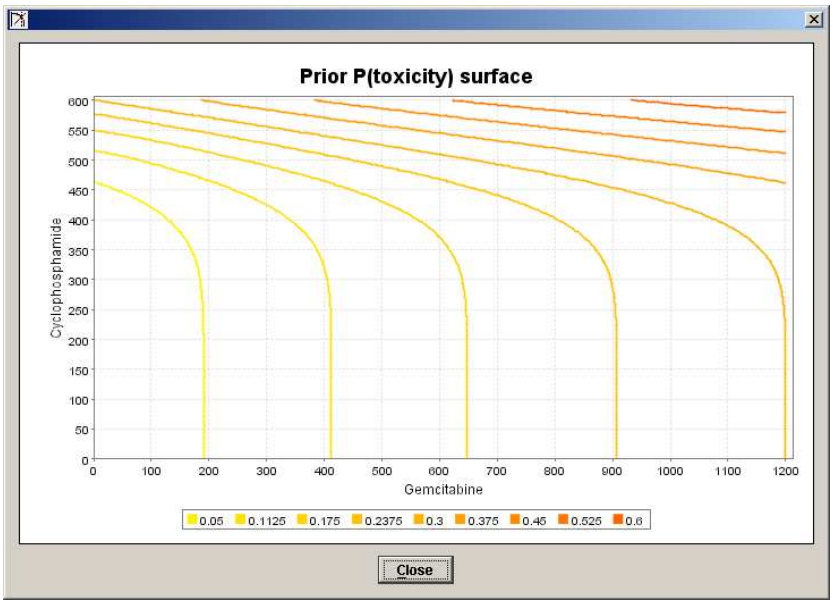
There is a limiting relation between $d^{(1)}$ and $d^{(3)}$ at which point the variance is zero, beyond which a variance cannot be found.

## 5.3 A 'strong' prior

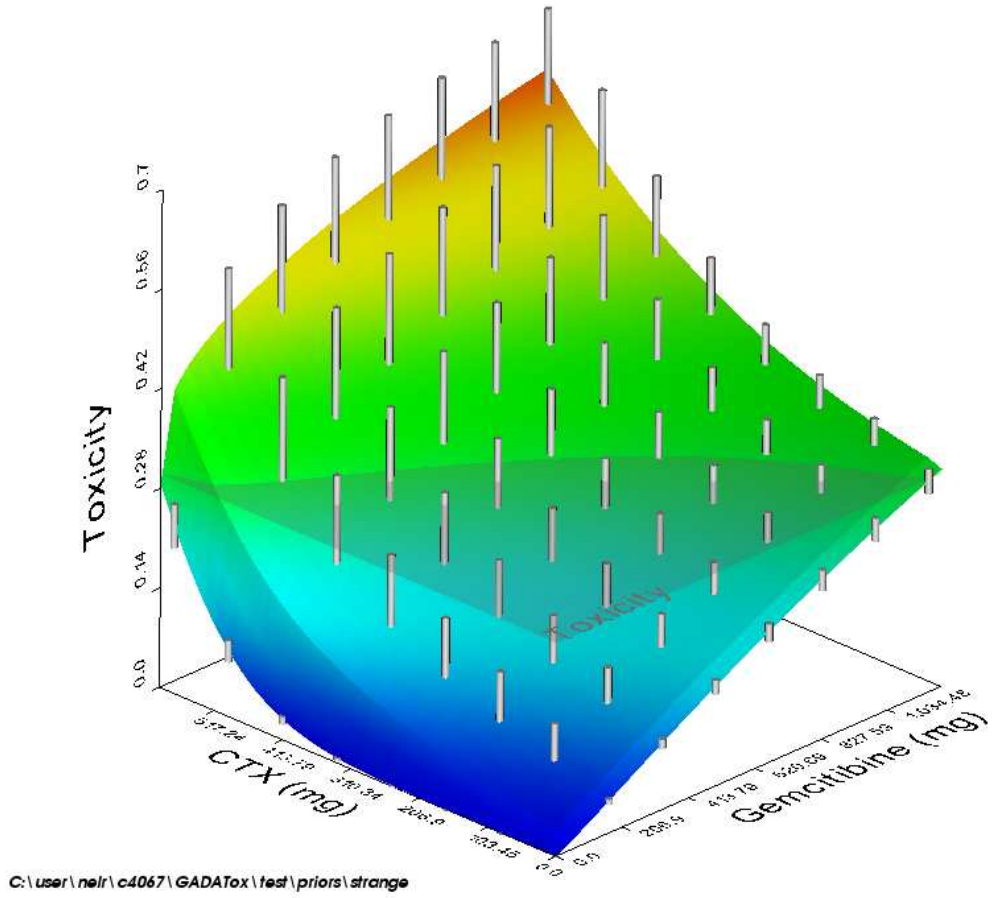
Specifying the following prior,

| | Gemcitabine (mg) | Cyclophosphamide (mg) |
|---|---|---|
| Highest dose with negligible toxicity ($d^{(1)}$) : | 120 | 350 |
| Maximum permitted dose ($d^*$, $d^{(2)}$) : | 1200 | 600 |
| Dose thought to have the prohibitively high toxicity rate ($d^{(3)}$) : | 3600 | 700 |
| Smallest dose known to have a toxicity rate above target ($d^{(4)}$) : | 2000 | 800 |

results in the toxicity of drug 1 depending very slowly on dose.




If the true surface is specified as "standard" then this contradicts the prior. Simulating, results in the following surface:



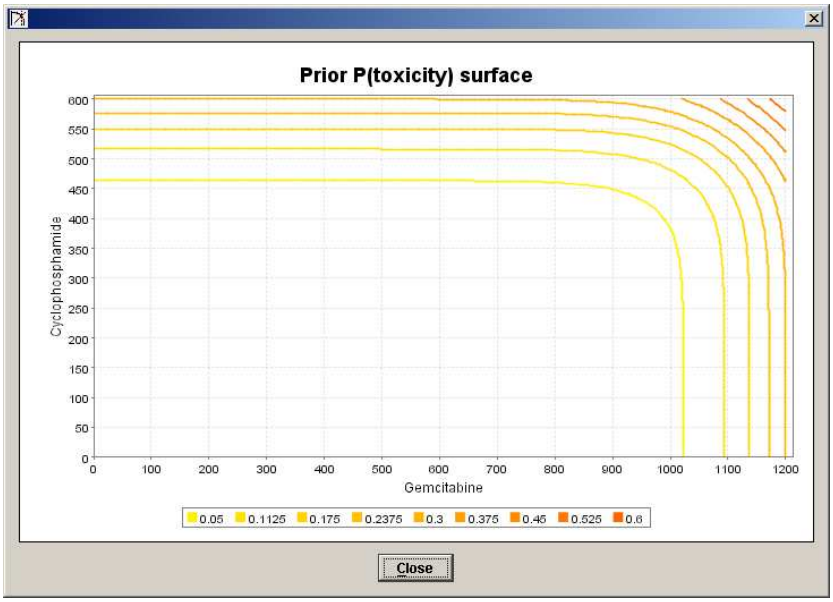

The surface is distorted showing that the prior is too strong and too incorrect to be overcome by the data. On the other hand the position of the 30% contour (where most of the subjects were placed) is approximately correct. The strange shape of the high and low toxicity regions is not important because those doses will never be used. Thus despite the inappropriate prior, the adaptive allocation has enabled us to find the answer we were looking for.

## 5.4 A 'weak' prior

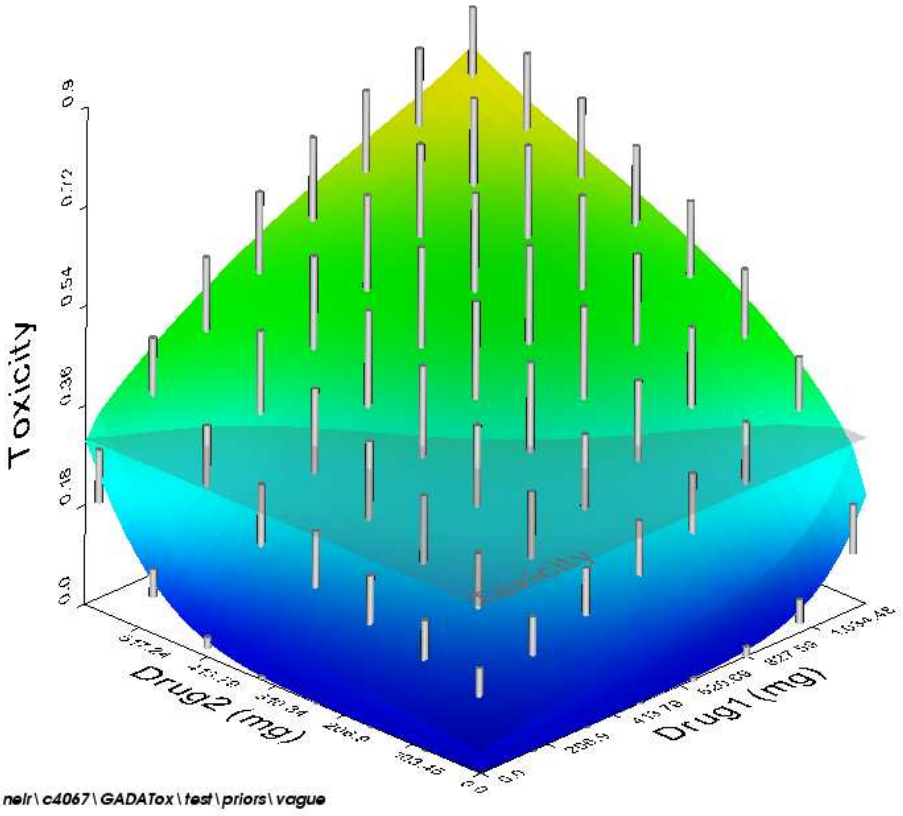
If instead we specify the following prior,

| | Gemcitabine (mg) | Cyclophosphamide (mg) |
|---|---|---|
| Highest dose with negligible toxicity ($d^{(1)}$): | 720 | 350 |
| Maximum permitted dose ($d^*$, $d^{(2)}$): | 1200 | 600 |
| Dose thought to have the prohibitively high toxicity rate ($d^{(3)}$): | 1320 | 700 |
| Smallest dose known to have a toxicity rate above target ($d^{(4)}$): | 1800 | 800 |

we get the following:




Simulating results for the standard scenario produces the following surface:




which is very close to the correct answer.