3-3-2005

Understanding the Continual Reassessment Method for Dose Finding Studies: An Overview for Non-Statisticians

Elizabeth Garrett-Mayer
Division of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, esg@jhu.edu

Understanding the Continual Reassessment Method for Dose Finding Studies: An Overview for Non-Statisticians

(Short title: Understanding the CRM)

Elizabeth Garrett-Mayer, PhD

Division of Oncology Biostatistics
Sidney Kimmel Comprehensive Cancer Center
Johns Hopkins University
Baltimore, MD 21205
esg@jhu.edu

Abstract:

The Continual Reassessment Method (CRM) has gained popularity since its proposal by O'Quigley et al. [1]. Many variations have been published and discussed in the statistical literature, but there has been little attention to making the design considerations accessible to non-statisticians. As a result, some clinicians or reviewers of clinical trials tend to be wary of the CRM due to safety concerns. This paper presents the CRM in a non-technical way, describing the original CRM with some of its modified versions. It also describes the specifications that define a CRM design, along with two simulated examples of CRMs for illustration.

Keywords: Continual reassessment method; clinical trials; phase I studies.

I. Introduction

The Continual Reassessment Method, developed by O'Quigley et al. [1], is a study design for dose finding (sometimes called Phase I) studies. Its proponents believe it to be superior to traditional dose-escalation designs because it "learns" from information gained at early time points in the study, it is less likely to treat patients at toxic doses and more likely to treat patients at efficacious doses. Many of its critics have concerns about its safety, worrying that escalation occurs too quickly and that patients may be treated at unsafe doses based on reliance on a mathematical model [2-4]. These concerns are largely unfounded. Current implementations of the CRM include criteria that disallow steep dose escalations and include the standard safety considerations of more traditional designs.

Many of the criticisms are in fact not criticisms that are specific to the CRM: they are applicable to most dose finding studies which treat 4 or fewer patients per dose level. Trialists often feel comfortable applying the standard "3+3" design (see Box 1), assuming that it is well-understood and safe [5]. However, the "3+3" design is in general no safer than the CRM or many other Phase I studies. Its main comfort-level stems from its common use. Most phase I studies tend to treat relatively few (i.e. 4 or fewer) patients at each of a small number of dose levels. As such, the certainty with which dose escalations, de-escalations, and the choices of dose levels are made tend to be poor. The "3+3" and other similar designs suffer from another flaw which is that users of this design often think that the chosen MTD is the dose which is most likely to have a toxicity rate of approximately 33%. This is not the case. In general, the true toxicity rate of the MTD in these studies will tend to be less than 33%, but toxicity rate estimation will be very imprecise given the few number of patients in the trial and at each dose level. However, in CRM-type designs, the toxicity rate is made explicit and the optimal dose (ergo, the MTD) is the dose that is estimated to be the one closest to achieving the desired toxicity rate.

One of the reasons that the CRM has not been more widely adopted is that it is not clearly understood. It has mathematical and statistical complexities that make it relatively difficult for many to understand. The goal of this paper is to explain the CRM, and some of its varieties, to the non-statistical reader. This does not imply that the CRM can be implemented without the consultation of a statistician: it is imperative that a statistician be involved in CRM trial design and CRM trial conduct. However, it is also imperative that (1) clinical investigators planning dose finding trials understand the CRM and how it might be an appropriate choice for their studies, and (2) members of institutional review boards and scientific review committees understand how the CRM works and compares to traditional designs in terms of safety and ethical considerations for patients. Note that in this paper, we focus on the dose-toxicity relationships. However, the CRM is more general, such that dose-efficacy relationships can also be explored using the CRM in a similar manner.

The paper is organized as follows. We provide some historical background of the CRM in section II. In section III, the steps of defining a CRM design are described. Section IV provides examples, including some practical issues in CRM trial conduct. The last section is a brief discussion.

II. Describing the CRM

II.a. The original CRM, introduced by O'Quigley et al. 1990

The CRM was first introduced by O'Quigley et al. and drew much attention from the biostatistical community [1]. The general idea behind the original CRM was that an a priori dose-toxicity curve

(DTC) was assumed and a desired toxicity rate was chosen. The estimated DTC was updated after each patients' toxicity outcome was observed, so that each patient's dose was based on information about how previous patients tolerated the treatment. For example, in figure 1, an a priori DTC is shown by the solid line where we can see that at the lowest dose level (dose level = 1), the probability of a dose-limiting toxicity (i.e. , DLT) is very close to 0. As dose level increases, the chance of DLTs increases. If we assume that the desired level of toxicity is 0.40 (meaning that it would be acceptable if 40% of patients had a DLT), then, according to this dose-toxicity curve, dose level 4 would be the optimal dose. So, assume that the solid curve in figure 1 is chosen a priori to represent our best guess at the dose-toxicity relationship. Then, the first patient in the trial would receive dose level 4 and would be observed for a DLT. However, if we were instead to assume that the dashed curve were our a priori guess about the dose toxicity curve, we would be more inclined to choose dose level 2 as the starting dose (i.e. the dose associated with 40% toxicity). An important point about the dose-toxicity curves shown in figure 1 is that they are based on mathematical models. Specifically, the curves shown in figure 1 are logistic curves which take values between zero and one and are monotonically related to dose. The logistic curves shown in figure 1 are "two-parameter" logistic models: each has a slope and scaling constant (like an intercept) which define it.

After the first patient has been treated at the dose chosen using the a priori DTC, whether or not s/he experienced a DLT is recorded. This information is then "combined" with the a priori dose-toxicity curve to get a better estimate of the true curve. In other words, we are updating our knowledge about the dose-toxicity relationship using the data that we accumulated from the first patient. The specifics of how this updating occurs will be described later in section IV. When we update our curve, we get a new estimated dose response curve. It likely looks much like our a priori curve, but is slightly shifted up or down depending on whether or not the first patient experienced a DLT. Updating with just one datapoint might not seem like much information, but as the number of patients accumulates, the dose-toxicity curve becomes estimated almost exclusively from the observed data and may look very little like the a priori curve that we began with. How much the curve changes will depend on how much weight we want to put on our a priori curve.

For those familiar with Bayesian statistics, the a priori curve represents our prior estimate of dose response and the accumulating information from patients represents our likelihood function. The resulting estimate of the dose-toxicity curve is our posterior estimate. The strength of the prior can be determined based on the precision parameters in the prior distributions chosen for the slope and scaling constant.

Using our updated dose-toxicity curve, we find what is now our best estimate of the optimal dose. By our earlier definition, this is the dose associated with a toxicity rate of 0.40. So, as before, we can graph our updated dose-toxicity model and find the dose that is associated with an estimated toxicity rate of 0.40. On figure 1, the dotted lines represent the two ways that we could have updated our curve: (1) the first patient could have experienced a DLT, in which case our updated DTC would fall slightly to the left of the solid a priori curve; or (2) the first patient could have not had a DLT, in which case the updated DTC falls slightly to the right of the a priori curve. If the former had occurred, the next patient would be treated at a dose of 3.9 and if the latter occurred at a dose of 4.2. The trial continues in this fashion: after each patient is treated, we re-estimate the DTC and treat the next patient at our new best estimate of the optimal dose. Patients continue to be treated until some level of certainty is achieved (stopping rules will be discussed in more detail in the next section). This is the general paradigm of O'Quigley et al's CRM.

This design grew criticism because of safety concerns. First, it was felt that large dose escalation steps might occur, subjecting individual patients to doses that were likely to cause DLTs. Second, beginning the CRM at the level deemed most appropriate by the a priori curve seemed to put patients in danger of exposure to toxic levels of the treatment: in many cases, the a priori curve was chosen with great uncertainty. The first of these criticisms of the original CRM has some merit. However, it can be argued that the second concern is not reasonable when comparing the CRM to other phase I designs which choose the initial dose level in much the same fashion as the CRM. Even so, as a result of both these concerns, several years later, modified CRMs were introduced.

Box 1: The standard "3+3" dose escalation design.

A. At each dose level, treat 3 patients beginning with dose level 1. Escalate to the next dose level or de-escalate to the previous dose according to the following rules:
   1. If 0 of 3 patients have a dose limiting toxicity (DLT), increase dose to next level.
   2. If 2 or more patients has a DLT, decrease dose to previous level.*
   3. If 1 of 3 patients has a DLT, treat 3 more patients at current dose level.
      a. If 1 of 6 has DLT, increase to next dose level.
      b. If 2 or more of 6 have DLT, decrease to previous level.
   4. If a dose has de-escalated to previous level:
      a. If only 3 had been treated at the previous level, enroll 3 more patients.
      b. If 6 have already been treated at the previous level, stop study and declare it the MTD.

B. The maximum tolerated dose (MTD) is defined as the largest dose for which 1 or fewer DLTs occurred.

C. Escalation never occurs to a dose at which 2 or more DLTs have already occurred.

* if de-escalation occurs at the first dose level, then the study should be discontinued.

II.b. Revised forms of the CRM

In the mid-1990's several variants of the CRM were introduced, addressing some of the safety concerns of the O'Quigley et al.'s proposed design. Goodman et al. retained the same general ideas of the CRM, with some modifications, calling the revised design the "Modified CRM" (see Box 2) [2]. The general idea is to begin the study like a "3+3" (or similar variant like a "2+2") and then switch to a CRM-type escalation procedure once the first DLT is observed. As such, the modified CRM requires a set of pre-specified doses in which to escalate. Once the CRM "kicks in," the modified CRM has some other key differences from the O'Quigley's version. First, the mathematical model is not solely responsible for determining the dose increases: dose increases are restricted by the pre-defined dose levels (i.e., large dose escalations are not permitted). A second main difference is that the modified CRM generally employs cohorts of size 2 or 3, unlike O'Quigley et al.'s size of 1. In practice, when there are no DLTs seen in a CRM design, it actually progresses much like the standard '3+3'. Additionally, when using the modified CRM, one of the major differences is that we do not use the a priori dose curve to choose our initial dose level: we use the lowest dose level under consideration.

Box 2: Goodman et al.'s "Modified CRM"

1. Pre-define dose levels for escalation as if for a "3+3" design.
2. Always start at the lowest dose level under consideration (i.e. in Figure 1, this would be dose level 1).
3. Enroll 1, 2, or 3 patients at each prescribed cohort (as opposed to only 1 per cohort).

4. Proceed as a standard dose escalation design in the absence of DLTs
5. Any given dose escalation cannot increase by more than one level, although dose de-escalations can be large.
6. Stop at a fixed a priori sample size.

Another variant was proposed at the same time by Möller [6]. Her motivation for revising the CRM stemmed from the idea that not a large enough range of doses tend to be considered in standard dose escalation designs, and so her design explores many doses and starts at a low dose level. Möller proposed two different revised CRMs. The first is the 'extended' CRM, where the design is split into two stages. In the first stage, the standard phase I '3+3' type of approach is used until one DLT is observed. Once a DLT is observed, the study enters the second stage where the design switches to a CRM where all of the information from stage 1 is included for estimating the DTC. In practice, this is very similar to Goodman et al.'s design described above. Möller's second proposed revision (the 'restricted' CRM) simply prevented escalations from exceeding one dose level—another facet of Goodman et al.'s modified CRM. The conclusions that Möller drew in evaluating these designs (i.e., the original CRM by O'Quigley, the restricted, and the extended) were that the original CRM tended to treat too many patients at high doses, the restricted CRM took many more patients to complete than the other two designs, but that all three provided unbiased estimates of the MTD.

Piantadosi et al. [7] proposed yet another similar variant of the CRM, which we will call the 'practical' CRM. The differences in Piantadosi et al.'s version are described in detail in Box 3.

Box 3: Piantadosi et al.'s 'practical' CRM.

1. Study preclinical information about toxicities.
2. Quantify clinical intuition about dose-toxicity by choosing dose level that would be guessed to incur low (e.g. 10%) toxicity and level that would be guessed to incur high (e.g. 90%) toxicity rate.
3. Estimate/draw the DTC assuming that the DTC passes through the dose-toxicity points described in 2. above.
4. Use the DTC curve to find the target dose for the desired toxicity level.
5. Treat 3 (or more) patients at the target dose.
6. Re-estimate the DTC based on the toxicity outcomes of the treated patients.
7. Revise estimate of the dose associated with high toxicity (from step 2 above).
8. Repeat steps 3-6 until target dose changes by less than 10% or meets another appropriate criterion for stopping.
9. Use target dose for future trials.

Because this is paper is not meant to describe every variant of the CRM that has been proposed, we refer the reader to other sources for more description of other CRMs [5, 8-13].

III. Defining a CRM design

There are two general properties that we commonly assume when defining CRM trials. We begin with them and then discuss how to go about defining a CRM trial. First, we assume that the DLTs are binary: patients either experience a DLT or they do not. If a patient experiences a toxicity, it is coded using the value 1, and if he does not, it is coded using the value 0. Second, we assume that the DTC curve is an increasing function of dose. In other words, as dose increases, the probability of toxicity increases.

IIIa. Choosing a dose-response model

One of the most important and often the most difficult part of designing a CRM is choosing the "functional form" (i.e. mathematical model) of the dose toxicity curve. It is important that a trained statistician be involved in choosing the mathematical model. And, it is critical that the statistician and clinicians both look at plots similar to those shown in figure 2 to have a clear understanding of the mathematical model that is being proposed. Reviewers of these trials might want to require visual displays of the proposed model—otherwise it can be difficult to discern the assumed relationship between dose and toxicity. There are a variety of models to choose from, but there are relatively few that tend to be used in practice and they fall into two categories: one parameter models, and two-parameter models. One parameter models, as their name implies, are somewhat simpler yet more restrictive than two-parameter models. The benefit of using one parameter models is that they require less information (i.e. fewer patients) to get a more precise estimate of the DTC. However, because they are more restrictive, they may not be flexible enough to accurately depict the true dose-toxicity relationship.

As part of the design of a CRM, we need to choose a model a priori which we deem consistent with our expectation of the relationship between dose and toxicity. Examples of four different functional forms are shown in Figure 2. The DTCs shown in figure 2A are based on the hyberbolic tangent function, as suggested by O'Quigley et al. [1]. Technically this means that,

$$p = \left(\frac{\tanh(d) + 1)}{2} \right)^{\alpha}$$

where p is the probability of toxicity and d is the dose. The unknown parameter of interest in the model is α (i.e., Greek letter "alpha"). In figure 2A, the curves each have a different value of α: the smaller the α, the higher the DTC in this case. When choosing the functional form for our CRM, we do not need to determine what we think α will be: just that we think that a curve from this "family" of curves is an appropriate way to model the relationship between dose and toxicity. The parameter α is estimated as part of the CRM: we use data about toxicity seen during the trial to estimate which α is most appropriate. In the case of figure 2A, if the toxicity associated with the treatment is relatively high, a small value of α will be estimated. However, if the toxicity is low, then the estimated value of α will be large.

The hyperbolic tangent model is just one way that we can quantify the relationship between dose and toxicity. Another common model is a one-parameter logistic model as used by Goodman et al. [2], for which examples are shown in figures 2B and 2C. The parameter α in this model determines the slope (and height) of the curve: the flattest of the curves has the lowest value of α and as the value of α increases, the DTC curve becomes steeper. There are several different ways in which a one-parameter logistic model can be parameterized. One of the key design choices in the one-parameter logistic CRM is that a constraint must be imposed such that *there must be one dose level at which all of the curves under consideration (i.e. for all levels of α) have the same assumed probability of toxicity*. In figure 2B, we can see that all of the DTCs converge at a probability of toxicity at 0.05 at dose level 1. In figure 2c, we use a different parameterization (or different mapping of doses) so that all curves pass through the toxicity rate of 0.90 at dose level 6.5. For trials where the primary interest is in toxicity, we would be interested in models such as those seen in 2c because we are interested in doses with low probability of toxicity (i.e. <0.40): in figure 2c, there is quite a lot of variation in the DTCs in the 0.10 to 0.30 range. However, if efficacy were the goal then we would want to make sure that the curves we were considering had large variability for high probability values (for efficacy, we usually want to find doses with high probabilities): in this case, figure 2B would be more appealing due to its flexibility in the range of 0.7 to 1.0. In general, when using a one-parameter logistic model, we will want to choose the dose

at which the curves converge to be either a very high dose (for toxicity-related trials) or a very low dose (for efficacy-related trials). This is because we do not expect to be iterating at values near the convergence point where there is little flexibility in the model. For more details of how to parameterize the a one-parameter logistic model, see Goodman et al. [2].

The two parameter logistic model, as is used by Piantadosi et al. [7], is a more attractive choice than the one parameter logistic model because both the slope (α) and intercept are estimated as part of the CRM. We are not constrained to only consider models which all converge at one dose level: estimating the intercept as part of the model allows us this flexibility. As mentioned above, the estimate of the DTC curve based on the two parameter model may not be as precise for a given sample size, but the model is more robust than the one-parameter model. In other words, it is less subject to "model-misspecification" as compared to the one-parameter model. However, allowing for the flexibility of two parameters, the parameter estimates may be unstable. Some models based on the two-parameter logistic function are shown in figure 2D. Notice that the curves take a variety of shapes, with varying slopes and intercepts.

The models shown in Figure 2 should provide some options for choosing your functional form. An example of another mathematical model recently implemented that is more ad hoc can be found in Mathew et al. [14].

IIIb. The spacing of the doses

In figures 2A through D, the doses are labeled as dose levels 1 through 7. These are simply "names" for dose levels and the actual doses corresponding to these levels need to be chosen by the clinical investigator. For example, in many clinical trials the 'modified Fibonacci' sequence is used to choose doses, where the increments are 2, 1.67, 1.5, 1.4, 1.33, 1.33, …, 1.33. Assume that we define dose level 1 as 100mg. Using the modified Fibonacci sequence, the next dose levels would be 200mg (2x100), 330mg (200x1.67), 500mg (330x1.5), and so on. We could "map" these doses to the corresponding dose levels such as those in figure 2.

But, the modified Fibonacci is not the only choice and it is certainly not necessarily the correct choice. For example, instead of using the modified Fibonacci, it might be sensible to use a log scale for the horizontal axis (e.g. dose level 1 = 100mg, dose level 2 = 1000mg, dose level 3 = 10000mg, etc.) or to make dose increments a constant increase (e.g. dose level 1 = 100mg, dose level 2 = 200 mg, dose level 3 = 300 mg, etc.). The increments chosen should depend on the preclinical information.

The easiest and most common way to solve this problem is to determine how you would choose your doses if you were to do a standard '3+3' or similar design. Depending on which version of the CRM you are using (as described in section II), you will often need to do this anyway. In most cases, based on preclinical or other preliminary data, the clinical investigator develops a series of dose levels that increase in increments in a predictable fashion (e.g. modified Fibonacci, log increases, etc.). These dose levels can then be mapped to the horizontal axis of the DTC plot.

IIIc. Number of patients per dose

As is stated above, the original CRM uses just one patient per dose level. However, it is quite common to see versions of CRMs implemented with more patients per dose level, given concerns over escalating too quickly. The number of patients per dose should depend on several factors, including (a) how conservative you would like to be in terms of quickly escalating, (b) how many patients are feasible, and (c) how precise you would like to be in your estimation of the dose toxicity relationship. The more patients per dose, the more information you will have about the dose- toxicity relationship. But, obviously, a trial with more patients per dose will take longer to

complete and will be more costly. Despite the trade-off between the precision and sample size, the safety of the patients should be the over-riding concern. As such, with precision and sample size as secondary considerations, the number of patients per cohort should be largely determined with regard to safety.

IIId. Target toxicity or response rate

One of the key study design parameters is the target toxicity rate (or response level, if efficacy is the outcome). In oncology trials of chemotherapy or other treatments that are given over a relatively short time period but have the potential for serious possibly life-threatening side effects, the target rate is often set between 0.20 and 0.30. In situations where the treatment is given as a long-term regimen, the target toxicity rate is likely to be less (in the range of 0.10 to 0.20). Again, there is not a standard that is used: the investigator should carefully consider how serious, tolerable and treatable the toxicities are for the treatment of interest. If efficacy is the goal of the study, then clearly we would want to set a relatively high threshold for our target rate depending on the nature of the response desired. If we are looking for clinical response, the rate will depend on the disease being studied and might range anywhere from 0.30 to .80. If biologic or pharmacologic response is of interest, setting a higher efficacy rate might be more practical, such as a rate of 0.80 or 0.90. Setting an efficacy level as high as 1.0, while seemingly appealing, is not practical and should generally not be used.

IIIe. Prior or design in absence of toxicities

In the original formulation of the CRM and in some of the modified versions, an a priori (or "prior") DTC was assumed. As is mentioned above, this a priori assumption is combined with the observed toxicity data to determine the dose at the next cohort of patients. The a priori DTC is a critical component of the CRM: without some assumption of the shape of the DTC, when we have not observed any toxicities, we cannot make any guess about what dose would incur toxicities at the desired rate. Imagine that at dose levels 1, 2, and 3 we see zero toxicities in a total of nine patients (three per cohort). This tells us, based on the observed data, that we estimate 0% toxicity at each of the three cohorts. And, it tells us nothing about at what dose we would expect the DTC to increase, or how quickly we would expect it to increase. This is critical information for choosing the subsequent dose. In the CRM, we get some information when the increase is expected to occur (and how quickly) from the a priori DTC. In a case where we have seen no toxicities, without the a priori DTC, we are not able to estimate the dose for the next cohort.

As an alternative to using an a priori curve, Goodman et al. [2] suggested beginning by starting with a '3+3' design, where a fixed number of doses are chosen and each of the doses has a defined quantity. In this case, we follow the '3+3' design until the first DLT is observed. At this point, the design switches to a CRM. Since one DLT has been observed, we have enough information for an estimate of the DTC without needing a formal a priori DTC. One of the benefits of this approach is that it is much more natural to many clinical investigators who are comfortable with choosing a set of doses to consider in a standard phase I trial than defining an a priori curve. Additionally, in this approach, once a DLT occurs, only the observed data is used for escalating: the chosen doses are only used to limit the amount of escalation, whereas the a priori DTC information is included at every iteration of the CRM.

III. Stopping and sample size

There are several commonly used stopping rules for CRM designs. In many cases, a fixed number of patients is decided at the onset of the trial and the CRM continues until the total sample size is reached, regardless of the number of doses considered and the number of patients treated per dose. This is particularly appropriate when dose levels can be continuous as opposed to discrete. This might be the case when the treatment is given as an infusion where the amount of drug given

can be quantified to very small increments. This would not be the case when the treatment under consideration comes in pill form where the pills are of a fixed dosage and so doses can only be given in pill-sized increments . Another approach that is more appropriate for discrete dose settings is to stop the trial once a fixed number of patients have been treated at any dose. For example, assume that in a trial, cohorts of 2 are used and the CRM has seemed to "settle" around a particular dose: it has treated 10 patients at that dose, and it recommends the same dose again for the next cohort. At this point, the trial would end and that dose would be considered the MTD. This is a difficult endpoint to work with when you allow for continuous dose levels, but it can be done. For example, Piantadosi et al. suggest stopping with the target dose changes by less than 10% [7]. Ishizuka and Ohasi suggest another stopping rule that depends on the precision of the DLT curve [10]. More discussion of stopping rules for CRMs is provided by Zohar and Chevret [15] and O'Quigley and Reiner [16].

IV. Two examples of the CRM

A. Example 1. In this design, we are using cohorts of size 2 and have a target DLT rate of 0.25. As in Goodman's modified CRM, we will begin with a standard escalation design until toxicity is observed and then we will allow the CRM to "kick in." We will assume discrete doses such that doses can be given in 50mg increments. The probability model that we will use is the hyperbolic tangent where the true doses are mapped to values are shown in Table 1.

In this simulated example, no toxicities were seen in dose levels 1, 2, and 3. At dose level 4 (400mg), there was 1 toxicity seen out of two patients. At this point the CRM is implemented, as is shown in Table 2. Using all of the information collected up through this cohort of patients, we can find the α value that is most consistent the data observed. In this example, the estimated α for of our model based on the toxicity information observed is 1.78. With this estimated α of 1.78, the most appropriate dose (i.e. the one that is our best estimate of the dose for a DLT rate of 0.25) is 350mg (which is shown as the dose for cohort 5). The CRM continues in this way. At the dose of 350, no toxicities are seen and our estimated α is 2.24, suggesting in increase in dose to 450.

The CRM continues to iterate until our stopping rule is achieved. In this study, the stopping rule was defined such that the trial terminates when 10 patients have been treated at a particular dose and the next cohort of the CRM would give the same dose. In the example above, after the 11th iteration, 10 patients have been treated at a dose of 450mg, and (if the CRM were to continue), the chosen dose for the next cohort is 450mg.

A graphical display of the iterations of the study (i.e., cohorts 4 through 15) is shown in figure 3. In this example, the CRM quickly converges to the area between 350 and 550mg doses and remains there for all of the cohorts after the CRM part of the study begins. So, there is movement in cohorts 4 through 15 (as can be seen in the table above), but it is confined to a relatively narrow range as compared to the initial range of doses under consideration (50mg to 800mg). In cohorts 7 through 15, the plots in figure 2 focus only on range of doses 350 through 550 to demonstrate how the CRM iterates to find the CRM, but it is important to point out that all of the toxicity information that is accumulated from all cohorts is used to estimate the DTC at every iteration of the CRM.

Based on the observed data, we can tabulate the toxicities observed at each of the doses, as shown in table 3. At the chosen dose (450mg), our estimated rate of DLTs given the observed data is 23%.

Example 2: The second example uses the method described in Piantadosi et al. (1998), where initial estimates of doses that produce high and low rates of DLTs are used to choose the first dose. We will assume that at 200 mg and 3000mg the rates of DLT are 10% and 90%, respectively. Our target DLT rate is 0.30. A two-parameter logistic model will be used to describe the relationship between DLT rate and dose. Our stopping rule is such that we will stop the trial when a total of 30 patients have been treated. With three patients per cohort, we will perform ten iterations of the CRM. In this example, we will assume that doses need not be discrete: dose to the nearest 1 mg can be accommodated (this is common when the treatment of interest is given as an infusion or the pharmacy can make capsules with drug levels that are very precise). Several safety mechanisms are also put into place: dose escalation cannot escalate by more than 400mg. The data presented in this example are generated based on a true model of toxicity: that is, the DLT data has been generated from a particular DTC curve.

Using our initial estimates of 10% DLT rate at 200mg and 90% DLT rate at 3000mg, we can estimate the best fitting logistic model by choosing the slope (β) and d50 level such that the model passes through these two points. In Figure 4A, our initial estimate of the DTC curve is shown via a dotted line. Using our initial estimate of the DTC curve, we find that the initial dose of 1060 mg is most consistent with our DLT rate of 0.30. And so, the first 3 patients are entered at a dose of 1060mg. Of these three patients, one has a DLT. Using this information, we recalculate the DTC and estimate the dose that is predicted to cause DLTs in 30% of patients. This continues until 30 patients (i.e., ten cohorts) have been treated.

Consider Figure 4A, which shows our initial estimate of the DTC before observing any data (dotted line), the estimated DTC based on the DLTs of the 30 patients in the study (solid line), and the true DTC from which the DLTs were simulated. In figure 4A, we see that even though our initial estimate was not very accurate, our estimate of the DTC based on 30 patients is very accurate, falling almost directly on top of the true DTC. However, this does not always occur in practice. Figures 4B, 4C, and 4D show three other possible estimated DTCs. In figure 4B, we see that the CRM has predicted the correct dose (or very close to it), but the actual shape of the curve is not very consistent with the shape of the true DTC. Figures 4C and 4D tend to under and over-estimate the dose, respectively. This is included to remind the investigator that, just like other Phase I designs, the CRM is not perfect: it will not find the correct dose everytime. However, there are steps one can take to increase the likelihood that the correct dose (or a dose very close to the correct dose) will be chosen.

Table 4 shows how precise the design in example 2 is (i.e. a total of 30 patients with cohorts of size 3), assuming the true DTC as shown in Figure 4A compared to two similar trials. Note that based on the true DTC and the target DLT rate of 0.30, the ideal dose is 1656 mg. The numbers in the table are based on simulations of each design (1000 per design). For each simulation, the "recommended dose" was recorded. These recommended doses for each design were then summarized to describe how accurate each design is in choosing an appropriate dose. Notice that designs 2 and 3 differ from design 1 in that both are larger (sample sizes are 50 and 60, respectively, versus a sample size of only 30 in design 1). Design 2 has the same number of cohorts as design 1, but there are five patients treated at each cohort versus only three as in design 1. In design 3, cohorts size is the same as in design 1, but there are twice as many cohorts (twenty versus ten). By increasing the sample size, we see that we get dose estimates closer to the true dose. For example, in design 1, 80% of trials recommend a dose within 400mg of the ideal dose, whereas in designs 2 and 3, 89% and 91% of trials recommend a dose within 400mg of the ideal. Four-hundred mg might seem like a large difference, but in looking at the DLT rates, most of the trials in all of the designs do not tend to overestimate the dose significantly in terms of DLT rate. For design 1, only 9% of trials recommend doses with a greater than 40% DLT rate, and for the

larger trials, only about 5% recommend dose with DLT rate greater than 40%. It is extremely rare for any of these designs to recommend a dose that has a DLT rate greater than 50% or less than 10%. Based on table 4, we can conclude that precision of our estimate will increase if we are to choose a larger sample size (either by increasing the number of cohorts or by increasing the number of patients per cohort). However, as cannot be seen in this example, the amount of additional information that you will gain by increasing the sample size will depend in large part on the shape of the true DTC curve: here we are only considering one scenario.

V. Discussion

The CRM is a novel and statistically appealing design for choosing an appropriate dose from a Phase I study. Some of its benefits include the ability to choose the desired toxicity level, a quantification of the accuracy of the dose in relation to the desired toxicity level, and its efficiency for treating patients at doses near to the optimal dose. In the past, it has been confusing to many clinical investigators because of its reliance on mathematical formulas for determining dose levels during the trial. For those not familiar with maximizing likelihood functions, the CRM should not be out of reach: the principles which underlie it are logical and (to some extent) need not be complicated by a lack of understanding of probability models. The goal of this article was to make the CRM accessible to those interested in applying or understanding it and minimizing the reliance on mathematics. As a result, some features of the CRM have been glossed over: this is not for a lack of appreciation of the complexity of the model, but instead for the purpose of bringing the CRM to the attention of a broader audience than it may have previously appealed to.

References

1. O'Quigley, JO, M Pepe, and L Fisher, *Continual Reassessment Method: A Practical Design for Phase I Clinical Trials in Cancer*. Biometrics, 46:33-48, 1990.
2. Goodman, SN, ML Zahurak, and S Piantadosi, *Some Practical Improvements in the Continual Reassessment Method for Phase I Studies*. Statistics in Medicine, 14:1149-1161, 1995.
3. Korn, E, D Midthune, T Chen, L Rubinstein, MC Christian, and R Simon, *A comparison of two phase I trial designs*. Statistics in Medicine, 13(18):1799-806, 1994.
4. Ratain, M, R Mick, RL Schilsky, and M Siegler, *Statistical and Ethical Issues in the Design and Conduct of Phase I and II Clinical Trials of New Anticancer Agents*. Journal of the National Cancer Institute, 85:1637-1643, 1993.
5. Storer, BE, *An Evaluation of Phase I Clinical Trial Designs in the Continuous Dose-response Setting*. Statistics in Medicine, 20(16):2399-2408, 2001.
6. 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, 14:911-922, 1995.
7. Piantadosi, S, JD Fisher, and S Grossman, *Practical Implementation of a modified continual reassessment method for dose-finding trials*. Cancer Chemotherapeutics and Pharmacology, 41:429-436, 1998.
8. Faries, D, *Practical Modifications of the Continual Reassessment Method for Phase I Cancer Clinical Trials*. Journal of Biopharmaceutical Statistics, 4:147-164, 1994.
9. Heyd, JM and BP Carlin, *Adaptive design improvements in in the continual reassessment method for phase I studies*. Statistics in Medicine, 18:1307-1321, 1999.
10. Ishizuka, N and Y Ohashi, *The continual reassessment method and its applications: a Bayesian methodology for phase I cancer clinical trials*. Statistics in Medicine, 20:2661-2681, 2001.
11. Thall, P and K Russell, *A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials*. Biometrics, 54:251-264, 1998.
12. Zohar, S and S Chevret, *Phase I (or Phase II) Dose-Ranging Clinical Trials: Proposal of a Two-Stage Bayesian Design*. Journal of Biopharmaceutical Statistics, 13(1):87-101, 2003.

13. Piantadosi, S and G Liu, *Improved designs for phase I studies using pharmacokinetic measurements*. Statistics in Medicine, 15:1605-1618, 1996.
14. Mathew, P, P Thall, D Jones, C Perez, C Bucana, P Troncoso, S Kim, I Fidler, and C Logothetis, *Platelet-derived growth factor receptor inhibitor imatinib mesylate and docetaxel: a modular phase I trial in androgen-independent prostate cancer*. Journal of Clinical Oncology, 22(16):3323-3329, 2004.
15. Zohar, S and S Chevret, *The continual reassessment method: comparison of Bayesian stopping rules for dose-ranging studies*. Statistics in Medicine, 20:2827-2843, 2001.
16. O'Quigley, JO and E Reiner, *A stopping rule for the continual reassessment method*. Biometrika, 85:741-748, 1998.

Table 1: Predefined dose levels for example 1.

DoseLevel	Dose	Mapped value
1	50mg	-1.5
2	100mg	-1.0
3	200mg	-0.5
4	400mg	0
5	800mg	0.5

Table 2: Progression of example 1. Cohort sizes are two. One patient experiences DLT in cohort 4, at which point the CRM kicks in, the α parameter is estimated, and the doses for the remaining cohorts are determined by the CRM.

Cohort	**Dose**	**Outcomes (0=no DLT, 1=DLT)**	**Estimated α**
1	50	0 0	--
2	100	0 0	--
3	200	0 0	--
4	400	0 1	1.78
5	350	0 0	2.24
6	**450**	0 0	2.74
7	550	1 0	2.40
8	**450**	0 0	2.72
9	500	0 0	3.03
10	550	1 1	2.29
11	**450**	0 1	2.11
12	400	0 0	2.27
13	**450**	1 0	2.13
14	400	0 0	2.26
15	**450**	0 0	2.39
16	**450**	**STOP**	

Table 3: Summary of results of example 1.

	DOSE							
	50	100	200	350	400	**450**	500	550
No DLT	2	2	2	2	5	**8**	2	1
DLT	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (17%)	**2 (25%)**	0 (0%)	3 (75%)
Predicted Rate of DLTs	<1%	<1%	4%	14%	19%	**23%**	27%	32%

Table 4. Characteristics of three CRMs of varying size with the same true model of dose-toxicity. Results are based on 1000 simulated trials for each design.

	Design 1 (example 2)	Design 2	Design 3
Total sample size	30	50	60
Patients per cohort	3	5	3
Number of cohorts	10	10	20
% of trials with recommended dose within 250mg of true dose (1656mg)	58%	69%	72%
% of trials with recommended dose within 400mg of true dose (1656mg)	80%	89%	91%
% of trials with recommended dose DLT rate of >40%	9%	6%	5%
% of trials with recommended dose DLT rate of >50%	0.5%	0.4%	0.4%
% of trials with recommended dose DLT rate of <20%	13%	6%	6%
% of trials with recommended dose DLT rate of <10%	0%	0%	0%

Figure 1: Dose-toxicity curves. Solid line and dashed line represent two different choices for a priori DTCs. The dotted lines represent updates to the solid a priori DTC after patient data has been observed.

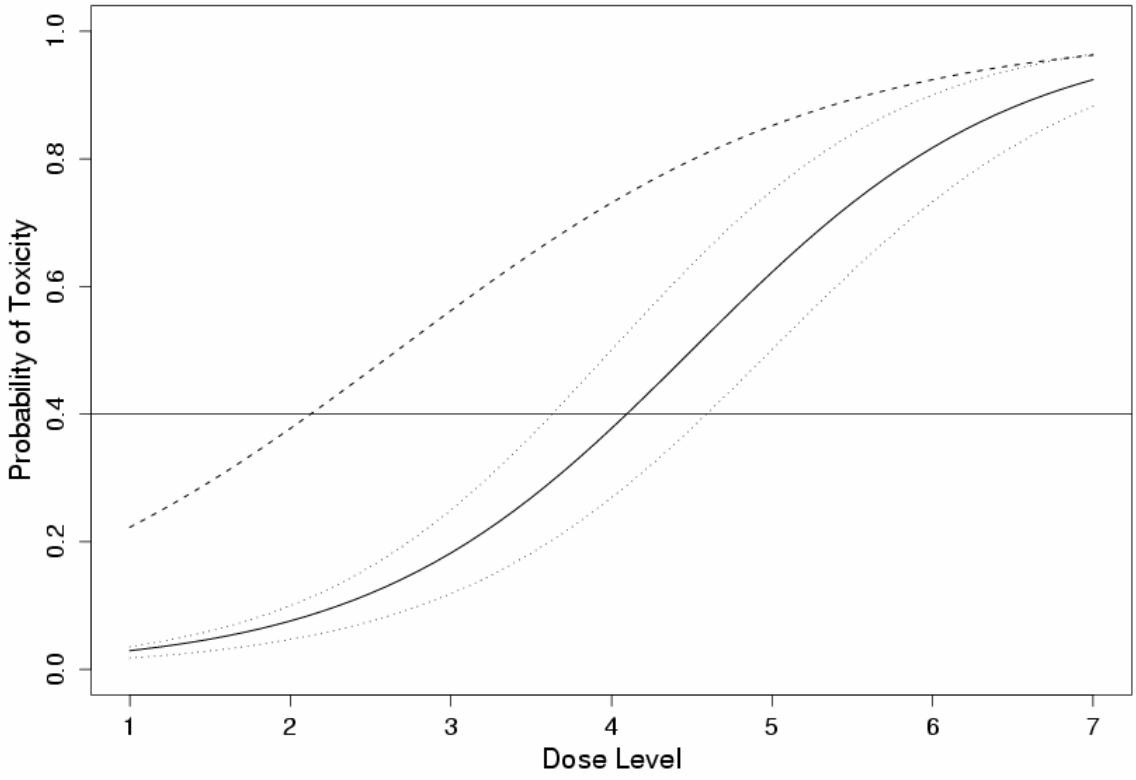


Figure 2: Four different mathematical models that could be used for the CRM. A: hyberbolic tangent model. B: one-parameter logistic model with constant value for d_5 (i.e. all curves impose the constraint that the probability of toxicity at dose level 1 is 0.05). C: one-parameter logistic model with different mapping of doses with constant value for d_{90} (such that all curves impose the constraint that at dose level 6.5 the probability of toxicity is 0.90). D: two-parameter logistic model.

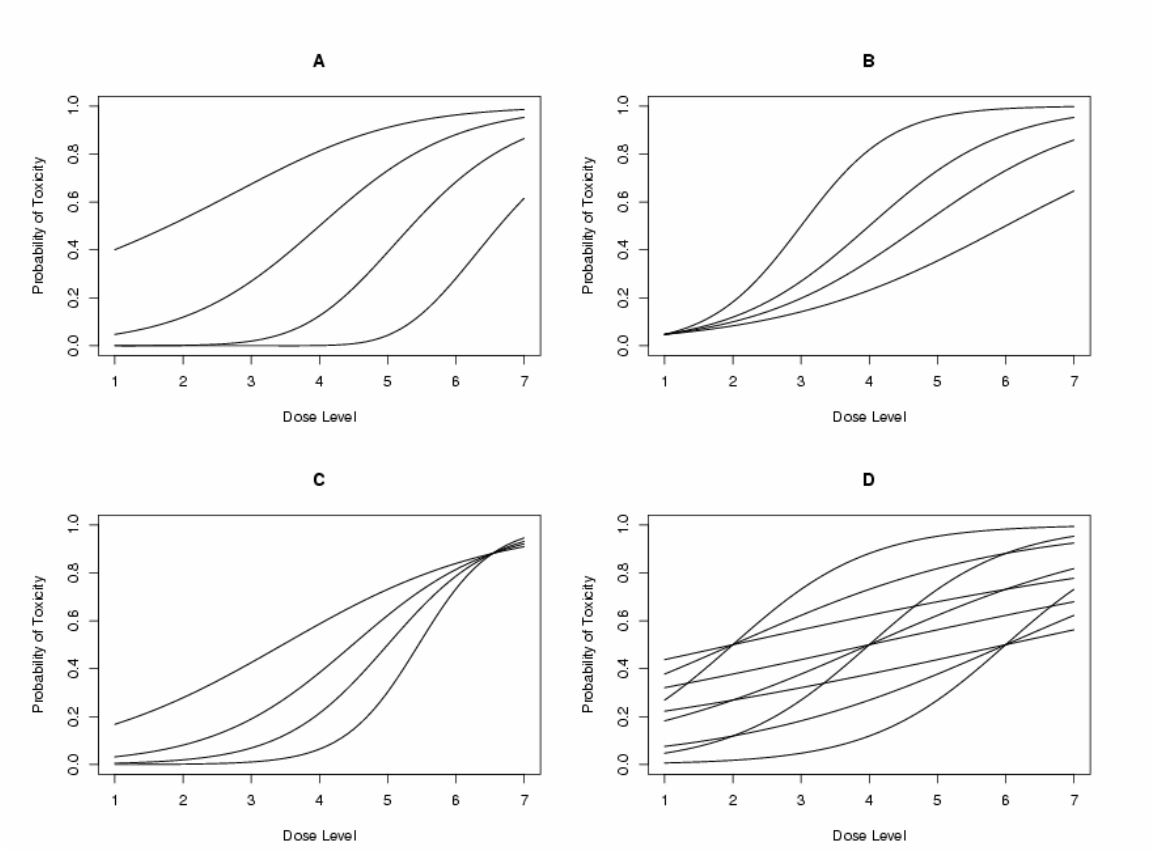


Figure 3: Demonstration of the CRM in example 1. At cohort 4, the first DLT occurs and the CRM kicks in. For each cohort, the accumulated toxicity information up to and including that cohort is used for estimating the value of α for the next cohort. In cohorts 7 through 15, the plots below focus on just the dose range of 350mg to 550mg to illustrated the movement of the modified CRM.

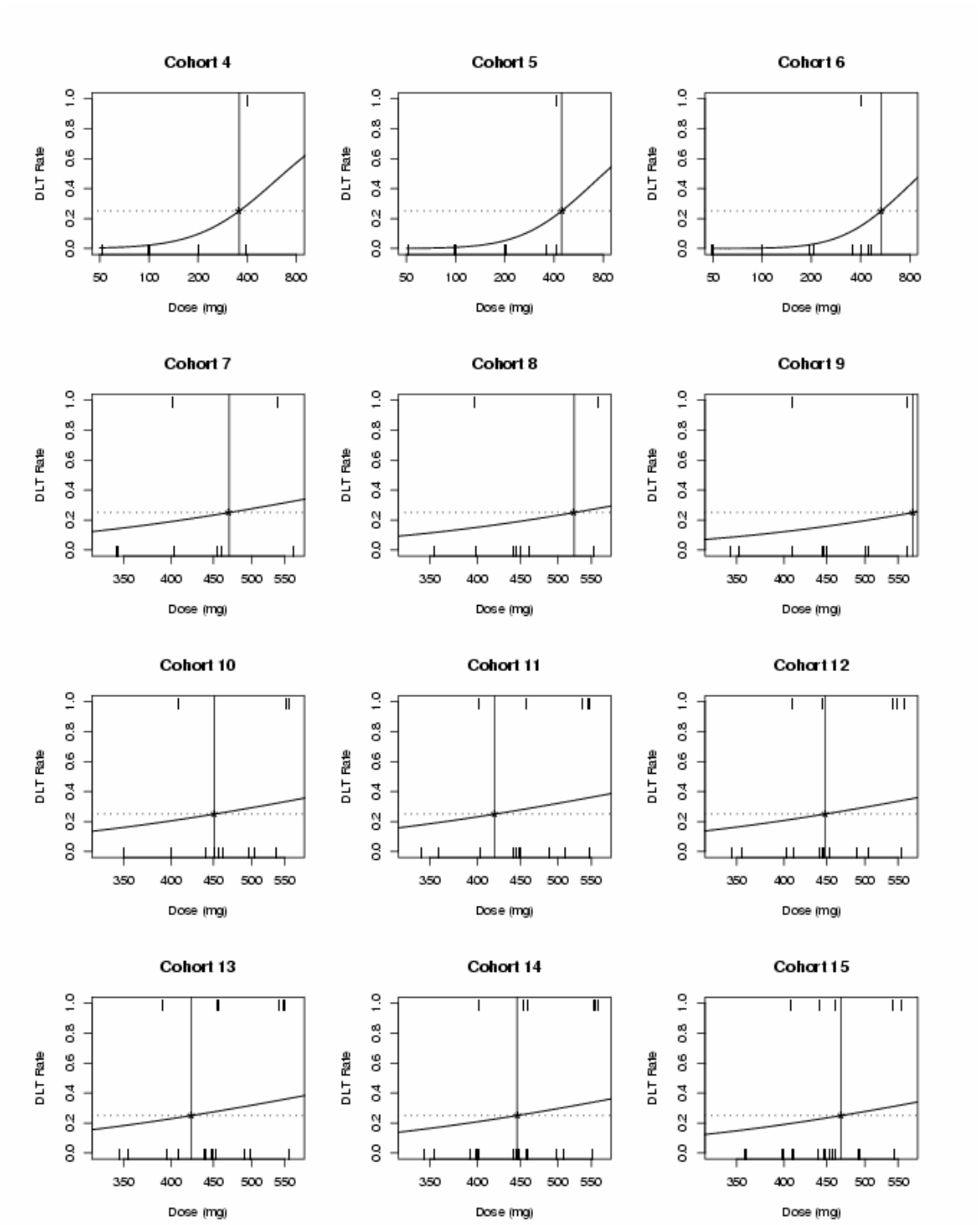
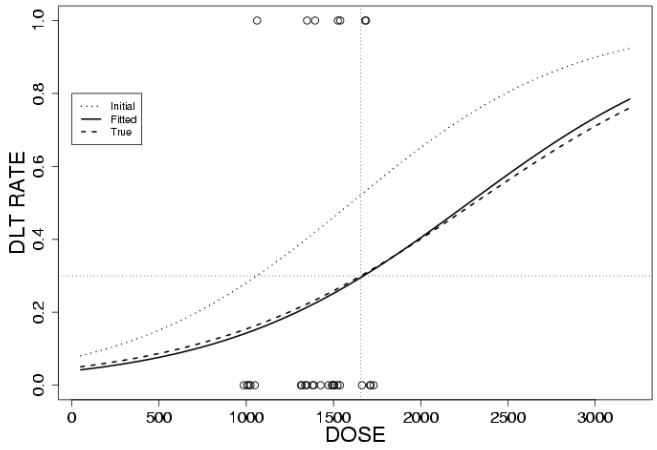
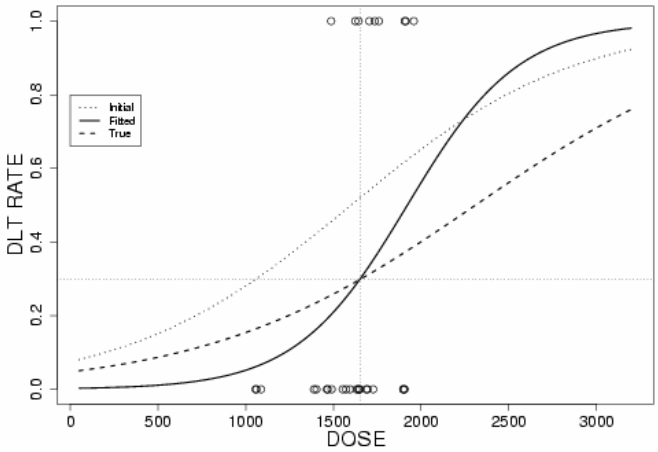


Figure 4. True DTC is dashed line, estimated DTC is solid line, and initial estimate of DTC is dotted line. A: Accurate estimation of the true DTC and dose, B: accurate estimation of dose, C: underestimate of dose and poor estimation of DTC, D: overestimate of dose and poor estimation of DTC.

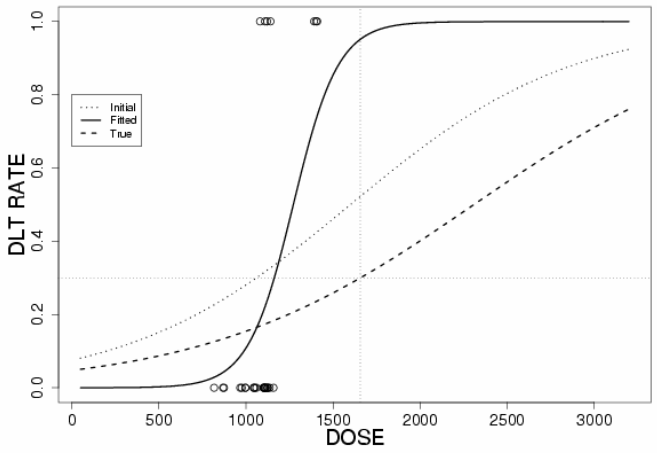
A



B



C



D