Published in final edited form as:

Stat Med. 2011 July 30; 30(17): 2117–2129. doi:10.1002/sim.4206.

Estimating the dose-toxicity curve in completed phase I studies

Alexia Iasonos* and Irina Ostrovnaya

Department of Epidemiology & Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

Summary

While there is an extensive amount of literature covering prospective designs for phase I trials, the methodology for analyzing these data is limited. Prospective designs select the maximum tolerated dose (MTD) through a dose-escalation scheme based on a model or empirical rules. For example, the “3+3” method (standard method: SM), assigns patients in cohorts of three and expands to six if one toxicity is observed. It has been shown previously that the MTD chosen by the SM may be low, possibly leading to a non-efficacious dose. Additionally, when deviation from the original trial design occurs, the rules for determining MTD might not be applicable. We hypothesize that a retrospective analysis would suggest a MTD that is more accurate than the one obtained by the SM. A weighted Continual Reassessment Method (CRM-w) has been suggested (O’Quigley 2005) for analyzing data obtained from designs other than the prospective Continual Reassessment Method (CRM). However, CRM-w has not been evaluated in trials that follow the SM design.

In this study, we propose a method to analyze completed phase I trials and possibly confirm or amend the recommended Phase II dose, based on a constrained maximum likelihood estimation (CMLE). A comparison of CRM, CRM-w, isotonic regression and CMLE in analyzing simulated SM trials shows that CMLE more accurately selects the true MTD than SM, and is better or comparable to CRM-w. Confidence intervals around the toxicity probabilities at each dose level are estimated using the cumulative toxicity data. A programming code is included.

Keywords

phase I trial; dose-toxicity curve; retrospective continual reassessment method; constrained maximum likelihood estimation; small sample size

1. Introduction

1.1 Motivation

The 3+3 dose escalation scheme, referred to as standard method (SM) [1], is the design most frequently used in current phase I trials [2], despite the fact that it often suggests a suboptimal dose level [3-4]. Since phase I trials are the first in-human trials, they typically start from a very low non-efficacious dose. The SM assigns patients in cohorts and prevents

*Corresponding Author: Alexia Iasonos, iasonosa@mskcc.org, 307 E. 63rd Str, 3rd floor, New York, NY 10065, Tel. 646 735 8115, Fax. 646 735 0010.
Both authors have contributed equally.

escalation of more than one dose level, therefore it tends to experiment at low dose levels. A sample size of six patients per dose level is not always attained in practice, since accrual stops when two patients experience a DLT. Hence, the SM often results in observed toxicity rates that are close or equal to zero at most dose levels and high at larger dose levels while the sample size per dose is very small. Consider as an example a trial that aimed to test six dose levels but escalation is stopped at the third dose level because two or more DLTs were observed. Table 1a shows six possible scenarios of such a trial. These scenarios consist of different observed rates, number of patients treated at each dose level and/or total sample size, however the recommended dose is the second level across these trials when using the SM. By definition, the observed toxicity rate at the MTD is zero or 0.167, but one can argue that dose level three is also acceptable in some of these scenarios. One of the major criticisms of this design is that it does not utilize the cumulative toxicity information across all dose levels and it does not target a pre-specified toxicity rate at the maximum tolerated dose (MTD) [5-6].

In this paper, we evaluate whether reanalysis of completed phase I trials that followed the SM would confirm or amend the choice of MTD. We propose an analysis method of Phase I data by estimating the dose-toxicity curve from completed trials. Examples where an estimated dose-toxicity curve can be useful are the following:

1. When phase I trials follow a deviated or modified plan of the SM in the dose escalation scheme which occurs very often in practice [2, 7]. In such cases, the prospective rule of the SM for determining the MTD might not be applicable, and a retrospective analysis method to estimate the MTD is needed.
2. When investigators amend an on-going trial to add additional dose levels because the visited levels were deemed too safe. They may wish to add intermediate levels between existing levels or new levels beyond the tested dose range.
3. When investigators want to switch from the SM to a sequential design such as the Continual Reassessment Method [8]. Accumulated data up to that point can be used to update the trial design parameters, and potentially provide an informative prior.

1.2 Background

Statistical literature deals with estimating the dose-toxicity curve in animal studies in particular. Some of the proposed approaches include: Parametric methods that assume a model of a probit or logit [9], or a sigmoid curve [10]; nonparametric methods that use isotonic regression [11-13]; or semi-parametric methods that use splines and kernels [14-16]. While these methods have been evaluated in animal experiments with sample sizes in the range of 50 per dose level, they cannot always be applied to phase I studies, which typically involve three to six patients per level and test very limited number of dose levels (e.g. 4-5). The proposed parametric approaches rely on model assumptions that cannot be verified by the data due to the small range in tested dose levels. The constraints involved in the sigmoid estimation [10] assume that a change point exists from concavity to convexity, which might not be true when available data are only observed at lower dose levels, i.e., the convex part of the curve that corresponds to the lower values of the dose space. Moreover, the observed toxicity rates in animal studies are usually much higher, while phase I studies in humans

often test low dose levels [5], and as a result, few toxicities are observed. For this reason, non-parametric methods, such as isotonic regression, may seem appealing in analyzing phase I data. When the observed toxicity rates increase monotonically with dose, the estimates from isotonic regression equal the observed rates [13]. In such cases, if the MTD from an isotonic regression is the dose level with an observed rate less than 0.33 then it will be the same as the MTD from the SM. However, if the observed rates do not increase monotonically or if the MTD is selected based on the proximity of the isotonic regression estimates to a pre-specified toxicity rate, then the two methods can result in different dose recommendation. This latter approach is evaluated in this manuscript.

Sequential designs, such as the Continual Reassessment Method (CRM), utilize likelihood or Bayesian framework to suggest a dose level based on data from accrued patients. One could use the same framework to analyze a completed trial that had followed any phase I design. He and colleagues [5] used parametric models within the CRM framework to estimate the MTD from trials that had followed the SM. They concluded that when the true toxicity rates follow the model, the method provides less biased estimates of the MTD compared with estimates by the SM; otherwise the method overestimates the expected toxicity levels. O'Quigley [17] argued that CRM cannot be used retrospectively as a method to analyze data gathered by another design, and he proposed the retrospective CRM instead. However, this method has not been validated in simulated trials or trials that followed the SM.

1.3 Objectives

In this paper, we propose a non-parametric method to re-analyze Phase I data by estimating the dose-toxicity curve from observed toxicity data of a completed phase I trial. We evaluate our proposed methodology as well as isotonic regression, CRM and retrospective CRM in simulated trials that followed the SM, trials with an expanded cohort at the MTD and existing trials from our institution (MSKCC). The outline of the paper is as follows: the methods are presented in Section 2 and evaluated via simulated trials in Section 3. In Section 4, we present an application in existing phase I trials. Section 5 presents confidence interval estimation and Section 6 concludes with a discussion.

2. Methods

2.1 Original likelihood-based CRM (CRM)

Using the same notation as O'Quigley [8], we assume the trial consists of k ordered dose levels, $d_1, d_2, \dots, d_k$, and a total of n patients. The visited dose level for patient j is denoted as x_j , and the binary toxicity outcome is denoted as y_j , where $y_j = 1$ indicates a dose-limiting toxicity (DLT) for patient j , and 0 indicates absence of a DLT. O'Quigley and Shen [18] used a one-parameter working model for the dose toxicity relation of the form, $\psi(d_i, a) = \beta_i^a$, where a is the unknown parameter, and β_i are the standardized units representing the discrete dose levels d_i . Since drugs are assumed to be more toxic at higher dose levels, $\psi(d_i, a)$ is an increasing function of d_i . The parameter estimate $\hat{a}$ can be obtained through a Bayesian framework [8] or maximum likelihood estimation [18].

We define t_j as the number of toxicities observed at dose level d_j and n_j the number of patients treated at dose level d_j among n patients, and let $\pi_j = n_j/n$ be the relative frequency with which d_j is allocated among n patients. The score function of the log-likelihood is given by:

$$U_n(a) = \sum_{i=1}^k \pi_i U_{in}(a), \quad (1)$$

where

$$U_{in}(a) = H(n_i) \left[\frac{t_i}{n_i} \frac{\psi'}{\psi}(d_i, a) + \left(1 - \frac{t_i}{n_i}\right) \frac{-\psi'}{1 - \psi}(d_i, a) \right], \quad i=1, \dots, k, \quad (2)$$

and $H(n_i) = I(n_i > 0)$, i.e. takes the value 1 when n_i is not equal to zero, and it is zero otherwise. The value $\hat{a}$ solves the equation $U_n(\hat{a}) = 0$. Once $\hat{a}$ is estimated by the current data, the MTD is the dose $d_0 \in \{d_1, \dots, d_k\}$ such that $|\psi(d_0, \hat{a}) - \theta^*|$ is minimized, where θ^* is a prespecified acceptable probability of toxicity (also known as target).

2.2 Weighted CRM (CRM-w)

When using CRM retrospectively, instead of solving (1), O'Quigley [17] suggested using a weighted estimating equation to find $\hat{a}$ when n patients have been observed. The rationale is the following: Had we used CRM from the start, the proportion of patients treated at the MTD would asymptotically tend to 1, zero elsewhere, as a result of CRM's ability to converge around a single dose level. Since the data are generated by a different design, we expect a larger dispersion in dose allocation than would be observed under a CRM design [19]. Thus, the weighted estimating equation is given by:

$$\sum_{i=1}^k v_i U_{in}(a) = 0, \quad (3)$$

The weights v_i are obtained by simulations and correspond to the percentage of patients CRM would have assigned to dose d_j . Specifically, a large number of CRM trials of size n are simulated using $\hat{R}_j = t_j/n_j$ as the true toxicity rates, where $\hat{R}_j$ are the observed toxicity rates from the existing trial. The weights v_i are estimated as the mean relative proportion of patients allocated by CRM at each dose level d_j across all simulated trials. O'Quigley showed that the final recommendation of MTD can change depending on whether equation (1) or (2) is solved for the unknown parameter a . However, when weighted CRM is applied in simulated trials, the following are possible limitations:

- a. The observed rates $\hat{R}_j = t_j/n_j$ are not necessarily monotonic, whereas CRM assumes monotonicity in the dose-toxicity curve. Smoothing these observed rates with isotonic regression, as previously suggested, forces $\hat{R}_j$ to be constant at various dose levels, and as a result, the weights based on the smoothed observed rates are not realistic, and the method's performance is not improved (data not shown). In addition, the observed toxicity rates are based on very few patients, and hence are crude estimates of the true toxicity rates. For example, there are trials where the

observed toxicity rates are too high early in the trial and lower at larger dose levels. Thus, the resulting estimated weights based on CRM trials are almost equal to 1 at the lowest dose level and equal to zero elsewhere, leading to no solution in equation (2) above.

- b. There are various versions of CRM designs that can be used in simulated trials that will result in different weights. The two-stage CRM described in O'Quigley and Shen [18], for example, can be used with one patient at a time or cohorts of three patients at each dose assignment until heterogeneity is observed among responses. The weights are more uniform across dose levels when cohorts of three patients are assigned at each dose level, since patients are allocated more evenly. In addition, if the sample size of the existing trial is small, simulated CRM trials with cohorts of three patients cannot be carried out without exceeding the observed sample size.

2.3 Constrained Maximum Likelihood Estimation (CMLE)

We propose a new methodology to estimate the dose-toxicity curve and hence recommend a new MTD based on CMLE. Since the observed toxicity rates t_i/n_i are based on small sample sizes, we developed a nonparametric method that is not restricted to a particular design. Instead, we impose linear constraints on the observed toxicity rates t_i/n_i , which include the assumption of monotonicity in the dose-response curve. The constraints control the increments in toxicity probabilities between adjacent dose levels. Specifically, we define the increment in toxicity rate between two dose levels as $\theta_i = p_i - p_{i-1}$, $i = 2, \dots, k$, where p_i is the probability of toxicity at d_i and let $\theta_1 = p_1$. Thus, $p_i = \sum_{j=1}^i \theta_j$. The estimated $\hat{p}_i$'s maximize the binomial log-likelihood:

$$\sum_{i=1}^k t_i \log(\sum_{j=1}^i \theta_j) + \sum_{i=1}^k (n_i - t_i) \log(1 - \sum_{j=1}^i \theta_j) \quad (4)$$

under the following constraints:

$$\delta_{\min} \leq \theta_i \leq \delta_{\max}, i=2, \dots, k, \text{ and } p_k = \sum_{j=1}^k \theta_j \leq 1, \quad (5)$$

where $\delta_{\min}$, $\delta_{\max}$ are two external parameters controlling the minimum and maximum increments in toxicity rates between any two consecutive doses. These parameters are chosen with the restriction that $\delta_{\min}, \delta_{\max} > 0$. The constraint in (5) guarantees that the dose-toxicity curve is increasing. If one assumes that a non-decreasing curve should be fit instead, setting $\delta_{\min} = 0$ would accommodate that, but herein we assume $\delta_{\min} > 0$. If $\delta_{\max} = 1$ and $\delta_{\min} = 0$, the likelihood is unconstrained and is maximized at $\hat{p}_i = t_i/n_i$. The solution of (4) under the constraints in (5) is obtained through numerical optimization [20]. Once $\hat{p}_i$'s are estimated, we select the MTD as the dose level closest to the target θ^* in Euclidean distance. We obtain the dose-toxicity curve by connecting the estimated $\hat{p}_i$'s.

Unlike CRM, CMLE is a nonparametric method that does not assume the shape of the dose-response curve is known. Instead, the number of patients treated at each dose level, the

number of observed toxicities and constraints are determining the estimated rates. The two constraint parameters, δ_{MIN} , δ_{MAX} , limit the flatness and steepness of the curve, respectively. Small values of δ_{MIN} lead to a conservative estimation by favoring a flatter curve, while larger values of δ_{MIN} impose larger jumps when the existing dose levels are low. Similarly, δ_{MAX} controls toxic dose levels by constraining the curve from increasing radically when high toxicity rates are observed. Note that there is no restriction on the toxicity rate at the first dose level, p_1 .

Investigators' clinical expertise should guide the choice of these two parameters with the rationale that the increase in toxicity is bounded within δ_{MIN} and δ_{MAX} . In the absence of clinical information about δ_{MIN} , δ_{MAX} in the specific trial, we suggest defining $\delta_{MIN} = \theta^*/k$, and $\delta_{MAX} = 1/(k-1)$, where θ^* is the target toxicity rate and k is the number of dose levels. Assuming that $p_1 = 0$, and no toxicities are observed, $\delta_{MIN} = \theta^*/k$ would allow the k^{th} dose to correspond to a toxicity rate of $\theta^*(k-1)/k$, which indicates a slowly increasing curve that has not reached its target rate by the k^{th} dose level. If $p_1 = 0$ and all patients have toxicities, a constraint of $\delta_{MAX} = 1/(k-1)$ would lead the k^{th} dose level to correspond to a toxicity rate of one, which results in the steepest curve. For example, Table 1b shows the predicted probabilities of toxicity using CMLE with the above constraints' values, where $\delta_{MIN} = 0.04$, $\delta_{MAX} = 0.20$, $k = 6$ and $\theta^* = 0.25$. The recommended dose level can vary from level two to three (indicated by *) depending on the observed toxicities. Even in the cases where the retrospective MTD is the same as the one recommended by the SM, ie level two, the estimated probabilities of toxicities vary to reflect the cumulative toxicities and the number of patients treated at each level. These estimated probabilities can be used to recommend new tested dose levels between existing visited levels when the trial is on-going or to amend the recommended phase II level. However these examples cannot show which method is closer to the true MTD since the true MTD is unknown. The simulations in the next section compare the accuracy of various methods relative to the true MTD.

3. Simulations

We evaluated the described methods in simulated trials generated under the SM design, since it is the most common design in practice. The dose-allocation scheme for the SM has been described by Korn *et al.* [1] and is provided in the appendix. A thousand simulated trials were generated using the SM, and the MTD was estimated retrospectively under the following schemes:

- Original design used to generate the data, i.e. MTD definition based on the SM
- Isotonic regression as described in section 2 of [13]
- Likelihood method as in CRM
- Retrospective or weighted CRM (CRM-w)
- Constrained Maximum Likelihood Estimation (CMLE).

Seven scenarios were evaluated under different true toxicity curves that allowed the location of the MTD to vary across the lower, middle, or high dose level. We also varied the number of dose levels from six (Scenarios S6.1-3) to five (Scenarios S5.1-2), and four (Scenarios

S4.1-2). The acceptable toxicity rate targeted at the MTD was fixed at 0.25. In CRM and CRM-w the power model, $\psi(d_i, a) = \beta_i^a$, was used where the standardized dose units are given by $\beta = (0.04, 0.07, 0.20, 0.35, 0.55, 0.70)$ for six dose levels as in O'Quigley [17]. Similarly, $\beta = (0.04, 0.07, 0.20, 0.35, 0.55)$ and $\beta = (0.04, 0.07, 0.20, 0.35)$ for five and four dose levels, respectively. Once the estimated toxicity probabilities $\hat{p}_i$'s are obtained via one of the methods above, the MTD is defined as the level with corresponding p_i closest to the target rate θ^* in Euclidean distance.

3.1 Simulations with true toxicity curves that meet the constraints

Table 2 shows the percentage of simulated trials that selected each dose level as the MTD for each method of analysis. The true toxicity rates used to simulate the data are shown in the second row, and the true MTD is highlighted in grey. For CMLE analysis, we used $\delta_{MIN} = \theta^*/k$, and $\delta_{MAX} = 1/(k-1)$, and all dose-toxicity curves satisfy these constraints. CMLE resulted in a more accurate estimate of the MTD compared with that provided by the SM, except for S6.1. In S6.1, isotonic regression performed better than the SM or CMLE, but the percent of trials selecting the MTD or the previous level was higher with CMLE (56%) compared with the SM (47%) or isotonic regression (53%). When the true MTD was among the last levels, and at least five levels were tested, CMLE was better than CRM or CRM-w. S5.2 is the only scenario where CMLE is slightly inferior than CRM-w. The explanation for the difference we observe between S5.2 and S4.2, lies in the fact that the constraint for δ_{MAX} is less restrictive in the latter example, allowing the estimated curve to be steep and thus closer to the true rates. Finally, note that in S5.1 and S4.1, the MTD is between the last two levels. In S4.1, CRM overestimates the toxicity rates at the last level, while CMLE more conservatively selects the dose level below the target toxicity.

3.2 Simulations with true toxicity curves that violate the constraints

We simulated trials from scenarios where the increments between adjacent dose levels violated the constraints and selected examples with six dose levels so that the constraints are violated at various places of the dose range, for both δ_{MIN} , δ_{MAX} values (Table 3). For a target of 0.25, $\delta_{MIN} = 0.04$ and $\delta_{MAX} = 0.20$. The three examples cover curves that are flat, steep, or sigmoid-type. These results show that CMLE is more accurate than all other methods when the MTD is among the lower dose levels, but its accuracy is lower at the first example when the MTD is the last dose level. These findings are consistent with the results of Table 2, showing that CMLE performs similarly, whether the constraints on the true underlying toxicity rates are satisfied or not.

3.3 Simulations with modified designs

Investigators are often interested in accruing more patients at the MTD in order to further validate the safety profile. As in Gonen [21], we simulated trials under the SM, followed by enrolling an additional 4 patients at the MTD for a total of 10. If 3 toxicities were observed out of 10, then the MTD remained the same. Otherwise, the method de-escalated and followed the 3+3 scheme by expanding a cohort to 6 patients until 2 DLTs out of 6 patients were observed. Expanded trials under scenarios 5.1, 5.2, 4.1, 4.2 are presented in Table 4. These results show that CMLE is more accurate than CRM-w when the true MTD is

at the second dose level, while the two methods are comparable when the true MTD is at the fourth or fifth dose level. The results across all scenarios illustrate that when the dose escalation scheme deviates from the 3+3 design by enrolling 10 patients at the MTD, all retrospective methods are more accurate than the SM by at least 10%. Had we allowed only 2/10 DLTs at the expanded cohort, the SM would deviate slightly further away from the true MTD and the retrospective analysis will be even more accurate.

4. Application in Phase I Completed Trials

We searched the MSKCC institutional database for phase I studies that closed between 1/1/2000 and 12/31/2006 with 20 patients accrued, and identified five protocols with toxicity data available and complete. The number of patients treated at each dose level deviated from the typical SM design, since investigators might have expanded a cohort to more than 6 patients or closed a cohort early. Figure 1 shows the five trials: Open circles indicate observed toxicity rates, while the black circle indicate the MTD recommended by the trial. The number of toxicities and patients treated at each dose level are shown on the upper axis. Details regarding each trial can be found in the following references [22-26]. Estimated toxicity rates were obtained by three methods: a) original CRM b) CRM-w, and c) CMLE. Isotonic regression estimates were the same as the observed rates, except for trial 99-024 and trial 99-083.

For example, trial 99-024 recommended the eighth dose; CRM-w, CRM, CMLE, and isotonic regression recommended the fourth, fifth, seventh and eighth dose, respectively. This trial illustrates unexpectedly high toxicity rates early in the trial, and observed rates that do not increase with dose. Simulating CRM trials that assume these crude rates as true, produces a steep dose-toxicity curve, and therefore CRM-w recommends a lower dose level as the MTD. CMLE depends less on the observed rates, instead it utilizes the cumulative information, and thus recommended a higher dose than the other two methods.

In trial 00-091, all methods agree with the dose selected by the SM. Trial 01-021 was analyzed in two versions: in version (a), five and four patients are treated at doses 400 mg and 600 mg, respectively, and the drug is given once a day. In version (b), the trial was amended to test 200 mg and 300 mg twice a day (bid), and patients treated at 200 bid and 300 bid were pooled with patients treated at 400 and 600 mg, respectively. In both versions, the trial recommended the second dose, i.e. 400 mg, as the MTD; whereas all three retrospective methods suggested the third dose (600 mg) based on the estimated curve.

Trial 99-083 is another example where all retrospective methods suggest one level above the original recommended dose of 120 mg, ie the fifth level of 155 mg. Finally, in trial 97-006, 14 dose levels are tested following an accelerated dose escalation plan assigning one patient per dose level followed by the SM. CMLE suggests dose 150, the same as that recommended by the trial, while both CRM and CRM-w suggest the last dose.

5. Confidence Interval Estimation

Thus far, we provided point estimates of the probabilities of toxicities, p_i . We now propose confidence interval (CI) estimation based on the constrained likelihood model in (4). First

we consider the upper limit and define u_i as the upper confidence limit for probability p_i , $i = 1, \dots, k$. A simple approach is to estimate the exact binomial limit [27] using information from the respective i^{th} dose level, i.e. $u_i: \Pr(T'_i \leq t_i) = \alpha/2$, where $T'_i \sim \text{Binomial}(u_i, n_i)$ and α denotes the confidence level. A more precise modification of this CI, is described by Wilson [27], and it is shown that the Wilson interval is conservative and wide given the small sample sizes typical in phase I trials. However, Morris [28] showed that the exact upper

binomial limit calculated using $(\sum_{j=i}^k t_j, \sum_{j=i}^k n_j)$ can also serve as the upper confidence limit u_i since $p_i \leq p_{i+1} \leq \dots \leq p_k$. This confidence limit might also be conservative if

$E(\sum_{j=i}^k t_j) > p_i$, however, pooling observations from several dose levels substantially increases the sample size and leads to a narrower interval. Morris [28] showed that a

weighted sum $s_i = \sum_{j=i}^k w_j t_j$ can be used to obtain an exact or conservative $(1 - \alpha)100\%$ upper confidence limit u_i for p_i in the following way:

u_i = upper limit of p_i in the constrained space, if $t_j = n_j$ for all $j = i, \dots, k$, or otherwise

$$u_i: \Pr(\sum_{j=i}^k w_j T'_j \leq s_i) = \alpha, \text{ where } T'_j \sim \text{Binomial}(u_i, n_j). \quad (6)$$

This approach works best if $p_i = p_{i+1} = \dots = p_k$. The weights w_j can be used to decrease the contributions from the dose levels with more toxicities when the differences among $p_i, \dots, p_k$ are large. An obvious selection of $w_j, j > i$, is $w_j = p_i/p_j$ since it satisfies $E(t_j) = E(w_j t_j)$. Assume $w_i = 1$. As the true probabilities are unknown, the weights can be approximated (shown in the appendix) using the constrained model given in (4) and they are equal to:

$$w_j \approx 1 - \frac{j-i}{j}. \quad (7)$$

The interval for the k^{th} dose, p_k , utilizes data only from the k^{th} dose. This approach is more efficient for lower dose levels, since it utilizes the information from toxicities from higher dose levels. Regardless of the weights, the CI we suggest is exact or conservative, as proved by Morris [28], and when the weights are close to p_i/p_j it is narrower compared with the one obtained when each dose level is considered separately.

In order to calculate the upper confidence limit, we modified the Morris method in (7)

similarly to Wilson's [27]. For a fixed i , note that $\sum_{j=i}^k w_j T'_j / \sum_{j=i}^k w_j n_j$ has an expectation equal to u_i and variance equal to $\sum w_j^2 n_j u_i (1 - u_i) / (\sum w_j n_j)^2$. To obtain u_i we solve:

$$u_i: \frac{\frac{\sum_{j=i}^k w_j t_j}{\sum_{j=i}^k w_j n_j} - u_i}{\sqrt{\sum w_j^2 n_j u_i (1 - u_i) / (\sum w_j n_j)^2}} = z_{\alpha/2}. \quad (8)$$

where $z_{\alpha/2}$ is $\alpha/2$ -percentile of the standard normal distribution. The lower confidence limit can be obtained as $1 - u_j$ if a success in the binomial distribution is defined as non-toxicity and the order of the dose levels is reversed.

For all the scenarios and simulated trials presented in Table 2, we compared the coverage and width of 95% CI obtained by the Wilson method given in (6), by the proposed CI using equal weights, and the weights given in (8). All 95% CIs described above are conservative and have 99-100% coverage for all scenarios. However, the width of the CI obtained via CMLE with the proposed weights is, on average, 3% smaller compared with equal weights, and 6% smaller compared with the Wilson binomial intervals obtained on individual dose levels separately (Table 5). Table 6 shows 95% CI for the MSKCC trials obtained via the three methods. Intervals based on the suggested weights are the narrowest compared with the other two methods, except for trial 99-024. The advantage of pooling information from subsequent dose levels to estimate the CI (last column of Table 6) is more apparent in trials with small number of patients per dose level, e.g. trial 97-006. However, when approximately 10 patients are treated per dose level, then the methods are comparable.

6. Discussion and Conclusion

Phase I trials determine the recommended dose level by prospectively following a dose-escalation algorithm based on a rule or a model. There is a plethora of Phase I trials that underestimate or select the wrong dose level either because the design is sub-optimal [6] or the design has not been followed stringently [2, 7]. In practice, the MTD properties resulting from the SM followed with modifications are unknown. Our retrospective analysis of simulated trials that have followed the SM algorithm confirmed that we can recommend a more accurate dose level without conducting a new trial. Hence, we provide an analysis method for Phase I data from a completed study with the goal of confirming or amending the MTD.

It is typical for Phase I trials to observe too low or high toxicity rates out of very few patients treated, as little as 1-2 patients per level. In fact, the number of patients treated at each dose level depends on the prospective design of the trial, as well as the true probability of toxicity at each level. However, for designs such as the SM, the allocation of patients to dose levels contains little information about the true toxicity rates, and the rule for determining the MTD does not use all cumulative toxicity data. Although, a retrospective analysis led to a more accurate estimate of the MTD for trials that followed the SM, we expect that the benefit of a retrospective analysis will be minimal for data obtained from probability-based and/or adaptive designs. Thus, we suggest performing a retrospective analysis of phase I trials for SM trials or trials that deviated from the initial design, since our simulations support a more accurate MTD in such cases.

We have proposed a non-parametric method based on constrained MLE (CMLE) to analyze Phase I trials retrospectively by fitting a dose-toxicity curve on the observed toxicity rates, and we have compared it to the retrospective weighted-CRM and isotonic regression. Based on the scenarios examined here and target rate of 0.25, isotonic regression performed better than the SM. Our simulations showed that <0.4% of the trials have observed rates that are

not in an increasing order, thus the isotonic regression estimates are expected to equal the observed rates in the majority of Phase I trials. In the context of retrospective analysis, CMLE does not solely rely on the observed rates on individual levels. Instead, it combines the cumulative toxicities with any prior information on the dose-toxicity curve through the constraints. Whereas CMLE is comparable or better than CRM-w, it is easier and faster (refer to R code in the appendix) to calculate, and a solution exists even for the trials where CRM-w fails to converge. Note that CRM depends on the standardized units which are selected to correspond to the anticipated toxicity probabilities at the respective dose levels. We have used the same units as in previous comparative studies [17-19], but simulations with other dose units (data not shown) can result in estimated toxicity rates that are different than the estimates provided here.

We have provided CI estimation based on the constrained model, and have evaluated these intervals in comparison to existing methods. A unique advantage of the proposed CI is that the interval for each dose level is taking into account the toxicity rates observed in subsequent levels. Thus, the interval provides the level of uncertainty in our estimates of toxicity rates utilizing the *cumulative data across levels* from the completed trial, an aspect that is not considered in the prospective design. We anticipate CI estimation to be more useful when new dose levels can be recommended by interpolating or extrapolating the estimated dose-toxicity curve within tested levels or outside the original tested range. However, further research is required in order to incorporate the actual dose units in the estimation method and potentially provide inferences to new dose levels.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Dr. Iasonos was funded by *Mr. William H. Goodwin and Mrs. Alice Goodwin and the Commonwealth Foundation for Cancer Research*, and *The Experimental Therapeutics Center of Memorial Sloan-Kettering Cancer Center*. The authors thank Drs. Colin Begg, Sujata Patil and Glenn Heller for their helpful comments and Carol Pearce for her review of this manuscript.

References

1. Korn E, Midthune D, Chen T, Rubinstein L, Christian M, Simon R. A comparison of two phase I trial designs. *Statistics in Medicine*. 1994; 13:1799–1806. [PubMed: 7997713]
2. Rogatko A, Schoeneck D, Jonas W, Tighiouart M, Khuri FR, Porter A. Translation of innovative designs into phase I trials. *J Clin Oncol*. 2007; 25:4982–4986. [PubMed: 17971597]
3. Ahn C. An evaluation of Phase I cancer clinical trial designs. *Statistic in Medicine*. 1998; 17:1537–1549.
4. Garrett-Mayer E. The continual reassessment method for dose-finding studies: a tutorial. *Clinical Trials*. 2006; 3:57–71. [PubMed: 16539090]
5. He W, Liu J, Binkowitz B, Quan H. A model-based approach in the estimation of the maximum tolerated dose in phase I cancer clinical trials. *Statistics in Medicine*. 2006; 25:2027–2042. [PubMed: 16025542]
6. Iasonos A, Wilton AS, Riedel ER, Seshan VE, Spriggs DR. A comprehensive comparison of the continual reassessment method to the standard 3+3 dose escalation scheme in Phase I dose-finding studies. *Clinical Trials*. 2008; 5:465–477. [PubMed: 18827039]

7. Zohar S, Lian Q, Levy V, Cheung K, Ivanova A, Chevret S. Quality assessment of phase I dose-finding cancer trials: proposal of a checklist. *Clin Trials*. 2008; 5:478–485. [PubMed: 18827040]
8. O'Quigley J, Pepe M, Fisher L. Continual Reassessment Method: A Practical Design for Phase I Clinical Trials in Cancer. *Biometrics*. 1990; 46:33–48. [PubMed: 2350571]
9. Prentice RL. A Generalization of the Probit and Logit Methods for Dose Response Curves. *Biometrics*. 1976; 32:761–768. [PubMed: 1009225]
10. Schmoyer RL. Sigmoidally Constrained Maximum Likelihood Estimation in Quantal Bioassay. *Journal of the American Statistical Association*. 1984; 79:448–453.
11. Stylianou M, Flournoy N. Dose finding using the biased coin up-and-down design and isotonic regression. *Biometrics*. 2002; 58:171–177. [PubMed: 11890313]
12. Stylianou M, Proschan M, Flournoy N. Estimating the probability of toxicity at the target dose following an up-and-down design. *Stat Med*. 2003; 22:535–543. [PubMed: 12590412]
13. Ivanova A, Flournoy N. Comparison of Isotonic Designs for Dose-Finding. *Statistics in Biopharmaceutical Research*. 2009; 1:101–107. [PubMed: 20157632]
14. Kong M, Eubank RL. Monotone Smoothing with Application to Dose-Response Curve. *Communications in Statistics - Simulation and Computation*. 2006; 35:991–1004.
15. Nottingham QJ, Birch JB. A Semiparametric approach to analysing dose-response data. *Statistics in Medicine*. 2000; 19:389–404. [PubMed: 10649304]
16. Staniswalis JG, Cooper V. Kernel Estimates of Dose Response. *Biometrics*. 1988; 44:1103–1119. [PubMed: 3233249]
17. O'Quigley J. Retrospective analysis of sequential dose-finding designs. *Biometrics*. 2005; 61:749–756. [PubMed: 16135026]
18. O'Quigley J, Shen L. Continual Reassessment Method: A Likelihood Approach. *Biometrics*. 1996; 52:673–684. [PubMed: 8672707]
19. O'Quigley J, Zohar S. Experimental designs for phase I and phase I/II dose-finding studies. *British Journal of Cancer*. 2006; 94:609–613. [PubMed: 16434987]
20. Byrd RH, Lu P, Nocedal J, Zhu C. A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing*. 1995; 16:1190–1208.
21. Gonen M. A Bayesian evaluation of enrolling additional patients at the maximum tolerated dose in Phase I trials. *Controlled Clinical Trials*. 2005; 26:131–140.
22. Azzoli C, Krug L, Gomez J, Miller V, Kris M, Ginsberg M, Henry R, Jones J, Tyson L, Dunne M, Pizzo B, Farmer A, Venkatraman E, Steffen R, Sirotinak F. A phase I study of pralatrexate in combination with paclitaxel or docetaxel in patients with advanced solid tumors. *Clin Cancer Res*. 2007; 13:2692–2698. [PubMed: 17473201]
23. Kelly W, O'Connor O, Krug L, Chiao J, Heaney M, Curley T, MacGregore-Cortelli B, Tong W, Secrist J, Schwartz L, Richardson S, Chu E, Olgac S, Marks P, Scher H, Richon V. Phase I study of an oral histone deacetylase inhibitor, suberoylanilide hydroxamic acid, in patients with advanced cancer. *J Clin Oncol*. 2005; 23:3923–3931. [PubMed: 15897550]
24. Kortmansky J, Shah M, Kaubisch A, Weyerbacher A, Yi S, Tong W, Sowers R, Gonen M, O'reilly E, Kemeny N, Ilson D, Saltz L, Maki R, Kelsen D, Schwartz G. Phase I trial of the cyclin-dependent kinase inhibitor and protein kinase C inhibitor 7-hydroxystaurosporine in combination with Fluorouracil in patients with advanced solid tumors. *J Clin Oncol*. 2005; 23:1875–1884. [PubMed: 15699481]
25. Krug L, Ng K, Kris M, Miller V, Tong W, Heelan R, Leon L, Leung D, Kelly J, Grant S, Sirotinak F. Phase I and pharmacokinetic study of 10-propargyl-10-deazaaminopterin, a new antifolate. *Clin Cancer Res*. 2000; 6:3493–3498. [PubMed: 10999734]
26. Shah M, Kortmansky J, Motwani M, Drobnjak M, Gonen M, Yi S, Weyerbacher A, Cordon-Cardo C, Lefkowitz R, Brenner B, O'Reilly E, Saltz L, Tong W, Kelsen D, Schwartz G. A phase I clinical trial of the sequential combination of irinotecan followed by flavopiridol. *Clin Cancer Res*. 2005; 11:3836–3845. [PubMed: 15897584]
27. Agresti A, Coull B. Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*. 1998; 52:8.
28. Morris MD. Small-Sample Confidence Limits for Parameters Under Inequality Constraints with Application to Quantal Bioassay. *Biometrics*. 1988; 44:1083–1092. [PubMed: 3233247]

Figure 1.

Estimated toxicity rates via three methods using data from existing Phase I trials. Open circles indicate observed toxicity rates; the black circle indicates the MTD selected by the trial. Trial 01-021 is presented in two versions (a) and (b) as described in Section 4.

Table 1a

Six examples of dose escalation with the standard method when level two is the recommended dose and escalation stopped at dose level three.

Number of DLTs/Number of patients treated		
Dose 1	Dose 2	Dose 3
0/3	0/6	2/3
0/3	0/6	2/6
0/3	1/6	2/3
0/3	1/6	2/6
0/3	1/6	4/6
1/6	1/6	4/6

Table 1b
Six examples of Table 1a where predicted probability is obtained by the retrospective analysis using constrained maximum likelihood estimation (CMLE)

	Dose 1	Dose 2	Dose 3
Number of DLTs/Number of patients treated	0/3	0/6	2/3
Predicted probability of toxicity	0	0.06	0.26
Recommended dose level by CMLE			*
Number of DLTs/Number of patients treated	0/3	0/6	2/6
Predicted probability of toxicity	0	0.04	0.24
Recommended dose level by CMLE			*
Number of DLTs/Number of patients treated	0/3	1/6	2/3
Predicted probability of toxicity	0	0.20	0.4
Recommended dose level by CMLE		*	
Number of DLTs/Number of patients treated	0/3	1/6	2/6
Predicted probability of toxicity	0	0.16	0.33
Recommended dose level by CMLE			*
Number of DLTs/Number of patients treated	0/3	1/6	4/6
Predicted probability of toxicity	0.04	0.24	0.44
Recommended dose level by CMLE		*	
Number of DLTs/Number of patients treated	1/6	1/6	4/6
Predicted probability of toxicity	0.17	0.30	0.50
Recommended dose level by CMLE		*	

* indicates the recommended level by CMLE.

Table 2

Percent of trials that selected each dose level out of 1000 simulated trials that followed the standard method (SM) as described in Appendix SA.2. Seven scenarios under true toxicity rates shown in the first row were evaluated using four methods: Isotonic regression (ISOTONIC); CRM; weighted CRM (CRM-w); and constrained maximum likelihood method (CMLE). Abbreviation: NF, not found.

S6.1: Dose Levels	0	1	2	3	4	5	6	NF
TRUE p_i		0.01	0.05	0.10	0.15	0.20	0.25	
SM	<1	3	10	17	22	21	26	
ISOTONIC		1	6	17	23	23	30	
CRM	<1	3	14	27	31	22	4	
CRM-w	1	2	14	28	29	22	4	
CMLE	<1	3	14	27	30	26		
S6.2	0	1	2	3	4	5	6	NF
TRUE p_i		0.05	0.10	0.15	0.20	0.25	0.35	
SM	3	9	20	20	19	20	9	
ISOTONIC	2	6	18	22	21	20	12	
CRM	<1	4	13	24	28	21	9	<1
CRM-w	1	5	11	25	28	22	8	<1
CMLE	<1	3	13	23	26	25	9	
S6.3	0	1	2	3	4	5	6	NF
TRUE p_i		0.05	0.10	0.25	0.35	0.55	0.70	
SM	3	9	41	30	15	3		
ISOTONIC	2	6	34	37	19	3		
CRM	<1	4	22	45	25	4		
CRM-w	1	5	21	44	27	3		
CMLE	<1	4	23	45	25	4		
S5.1	0	1	2	3	4	5	NF	
TRUE p_i		0.05	0.10	0.15	0.20	0.30		
SM	3	9	21	21	24	22		
ISOTONIC	2	6	18	23	27	25		

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

S6.1: Dose Levels	0	1	2	3	4	5	6	NF
CRM	<1	4	13	25	34	21	3	
CRM-w	1	5	11	25	33	22	3	
CMLE	<1	4	13	24	35	25		
S5.2	0	1	2	3	4	5	NF	
TRUE p_i		0.15	0.25	0.30	0.40	0.50		
SM	15	34	25	14	6	1	6	
ISOTONIC	11	29	30	17	6	1	6	
CRM	9	22	36	24	8	1		
CRM-w	11	24	32	23	9	1	<1	
CMLE	14	23	30	23	9	1		
S4.1	0	1	2	3	4	NF		
TRUE p_i		0.05	0.10	0.15	0.30			
SM	3	9	22	38	29			
ISOTONIC	2	6	19	37	37			
CRM	<1	4	13	34	44	5		
CRM-w	1	5	11	35	43	5		
CMLE	<1	3	14	46	37			
S4.2	0	1	2	3	4	NF		
TRUE p_i		0.15	0.25	0.30	0.40			
SM	15	34	26	14	6	6		
ISOTONIC	11	29	30	17	7	6		
CRM	9	22	36	23	10	<1		
CRM-w	10	24	32	23	10	<1		
CMLE	9	20	37	27	7			

Table 3

Percent of trials that selected each dose level out of 1000 simulated trials that followed the standard method (SM) as described in Appendix SA.2. Three scenarios under true toxicity rates that violate the constraints ($\delta_{MIN}=0.25/6=0.04$, $\delta_{MAX}=1/5=0.2$) were evaluated using four methods: Isotonic regression (ISOTONIC); CRM; weighted CRM (CRM-w); and constrained maximum likelihood method (CMLE). Abbreviation: NF, not found.

Case 1:						
Dose Levels	1	2	3	4	5	6 NF
TRUE p_i	0.0001	0.01	0.02	0.03	0.04	0.25
SM		<1	2	2	44	53
ISOTONIC		<1	<1	14	28	70
CRM			<1	2	17	68 12
CRM-w		<1	<1	2	25	60 12
CMLE			<1	2	45	53
Case 2:						
Dose Levels	1	2	3	4	5	6 NF
TRUE p_i	0.0001	0.01	0.25	0.30	0.40	0.45
SM		45	29	18	6	2
ISOTONIC		25	46	20	7	2
CRM		5	53	29	11	2 <1
CRM-w		6	53	29	11	2 <1
CMLE		5	54	30	10	2
Case 3:						
Dose Levels	1	2	3	4	5	6 NF
TRUE p_i	0.0001	0.01	0.04	0.25	0.30	0.33
SM		2	45	27	17	10
ISOTONIC			<1	28	41	18 12
CRM		0	3	60	26	11 <1
CRM-w		<1	10	51	25	13 <1
CMLE		0	11	59	22	9

Comparisons of the methods in expanded trials that accrued 10 patients at the MTD. Percent of trials that selected each dose level out of 1000 simulated trials that followed the standard method (SM) before and after expansion at the MTD. Four scenarios were evaluated corresponding to scenarios 5.1-4.2 of Table 2.

Table 4

S5.1	Dose Levels	0	1	2	3	4	5	NF
	TRUE p_i		0.05	0.10	0.15	0.20	0.30	
	SM	3	9	21	21	24	22	<1
	SM after expansion	3	12	25	21	23	17	<1
	ISOTONIC	1	5	16	23	29	25	<1
	CRM	<1	3	9	26	40	22	
S5.2	CRM-w	1	5	13	26	34	21	
	CMLE	<1	4	15	25	33	23	
	Dose Levels	0	1	2	3	4	5	NF
	TRUE p_i		0.15	0.25	0.30	0.40	0.50	
	SM	15	34	25	14	6	1	6
	SM after expansion	15	36	25	14	5	<1	6
S4.1	ISOTONIC	9	29	31	18	6	1	6
	CRM	11	22	30	28	8	1	<1
	CRM-w	14	27	30	22	7	<1	
	CMLE	13	26	35	19	6	1	
	Dose Levels	0	1	2	3	4	NF	
	TRUE p_i		0.05	0.10	0.15	0.30		
S4.2	SM	3	9	22	38	29	<1	
	SM after expansion	3	13	22	39	24	<1	
	ISOTONIC	1	5	16	37	41	<1	
	CRM	<1	3	9	35	52		
	CRM-w	1	5	14	37	42		
	CMLE	<1	4	16	42	38		
S4.2	Dose Levels	0	1	2	3	4	NF	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

SS.1	Dose Levels	0	1	2	3	4	5	NF
	TRUE p_i	0.15	34	26	0.30	0.40		
	SM	15	34	26	14	6	6	
	SM after expansion	15	37	25	14	4	6	
	ISOTONIC	9	29	32	18	6	6	
	CRM	11	22	30	28	9	<1	
	CRM-w	14	27	30	22	7		
	CMLE	13	26	35	20	6		

Table 5

Average width of 95% Confidence Intervals (CI) across all dose levels for each scenario with three methods: Wilson, CMLE with equal weights, and CMLE with weights given in Section 5.

Scenario	Wilson CI	CMLE with Equal weights CI	CMLE with Weighted CI
6.1	0.55	0.46	0.44
6.2	0.56	0.51	0.48
6.3	0.56	0.57	0.54
5.1	0.56	0.50	0.48
5.2	0.57	0.60	0.56
4.1	0.55	0.50	0.48
4.2	0.57	0.59	0.55

Table 6

95% Confidence Intervals for the existing trials based on Wilson's method; CMLE with equal weights; and CMLE with weights given in equation (8) of Section 5.

Dose Level	Number of DLT/Number of patients	Wilson Method	CMLE Equal Weights	CMLE Weighted CI
Trial 99-024		95% Confidence Interval		
1	0/3	(0.00, 0.56)	(0.00, 0.26)	(0.00, 0.28)
2	0/5	(0.00, 0.43)	(0.00, 0.28)	(0.00, 0.31)
3	2/6	(0.01, 0.70)	(0.04, 0.32)	(0.05, 0.38)
4	1/5	(0.01, 0.62)	(0.06, 0.29)	(0.06, 0.31)
5	0/3	(0.00, 0.56)	(0.05, 0.28)	(0.05, 0.29)
6	1/7	(0.01, 0.51)	(0.05, 0.33)	(0.05, 0.34)
7	0/3	(0.00, 0.56)	(0.05, 0.39)	(0.04, 0.39)
8	0/3	(0.00, 0.56)	(0.05, 0.56)	(0.02, 0.56)
Trial 00-091				
1	0/3	(0.00, 0.56)	(0.00, 0.42)	(0.00, 0.36)
2	0/3	(0.00, 0.56)	(0.00, 0.46)	(0.00, 0.42)
3	0/3	(0.00, 0.56)	(0.00, 0.52)	(0.00, 0.49)
4	3/12	(0.09, 0.53)	(0.05, 0.59)	(0.07, 0.57)
5	3/5	(0.23, 0.88)	(0.11, 0.88)	(0.17, 0.88)
Trial 01-021 (a)				
1	0/6	(0.00, 0.39)	(0.00, 0.40)	(0.00, 0.33)
2	0/5	(0.00, 0.43)	(0.00, 0.52)	(0.00, 0.47)
3	1/4	(0.01, 0.70)	(0.01, 0.69)	(0.02, 0.68)
4	3/6	(0.19, 0.81)	(0.08, 0.81)	(0.14, 0.81)
Trial 01-021 (b)				
1	0/6	(0.00, 0.39)	(0.00, 0.40)	(0.00, 0.32)
2	0/9	(0.00, 0.30)	(0.00, 0.48)	(0.00, 0.42)
3	4/10	(0.17, 0.69)	(0.06, 0.67)	(0.09, 0.66)
4	3/6	(0.19, 0.81)	(0.11, 0.81)	(0.17, 0.81)
Trial 99-083				
1	0/3	(0.00, 0.56)	(0.00, 0.37)	(0.00, 0.32)
2	0/3	(0.00, 0.56)	(0.00, 0.40)	(0.00, 0.35)
3	0/3	(0.00, 0.56)	(0.00, 0.43)	(0.00, 0.40)
4	0/9	(0.00, 0.30)	(0.00, 0.48)	(0.00, 0.45)
5	5/11	(0.21, 0.72)	(0.08, 0.67)	(0.12, 0.67)
6	2/5	(0.12, 0.77)	(0.10, 0.77)	(0.15, 0.77)
Trial 97-006		95% Confidence Interval		

Dose Level	Number of DLT/Number of patients	Wilson Method	CMLE Equal Weights	CMLE Weighted CI
1	0/1	(0.00, 0.95)	(0.00, 0.28)	(0.00, 0.33)
2	0/1	(0.00, 0.95)	(0.00, 0.29)	(0.00, 0.29)
3	0/1	(0.00, 0.95)	(0.00, 0.30)	(0.00, 0.28)
4	0/1	(0.00, 0.95)	(0.00, 0.31)	(0.00, 0.29)
5	0/1	(0.00, 0.95)	(0.00, 0.32)	(0.00, 0.30)
6	0/1	(0.00, 0.95)	(0.00, 0.33)	(0.00, 0.31)
7	0/1	(0.00, 0.95)	(0.00, 0.35)	(0.00, 0.32)
8	0/1	(0.00, 0.95)	(0.00, 0.36)	(0.00, 0.34)
9	0/1	(0.00, 0.95)	(0.00, 0.38)	(0.00, 0.36)
10	0/1	(0.00, 0.95)	(0.00, 0.39)	(0.00, 0.38)
11	0/3	(0.00, 0.56)	(0.00, 0.41)	(0.00, 0.39)
12	0/3	(0.00, 0.56)	(0.00, 0.48)	(0.00, 0.46)
13	1/9	(0.01, 0.43)	(0.01, 0.57)	(0.01, 0.56)
14	2/2	(0.34, 1.00)	(0.04, 1.00)	(0.12, 1.00)