HHS Public Access

Author manuscript

J Biopharm Stat. Author manuscript; available in PMC 2016 September 18.

Published in final edited form as:

J Biopharm Stat. 2015 ; 25(6): 1206–1214. doi:10.1080/10543406.2015.1086779.

Enrollment and stopping rules for managing toxicity requiring long follow-up in Phase II oncology trials

Guochen Song and
Quintiles

Anastasia Ivanova
Department of Biostatistics, University of North Carolina at Chapel

Abstract

Monitoring of toxicity is often conducted in Phase II trials in oncology to avoid an excessive number of toxicities if the wrong dose is chosen for Phase II. Existing stopping rules for toxicity use information from patients who have already completed follow-up. We describe a stopping rule that uses all available data to determine whether to stop for toxicity or not when follow-up for toxicity is long. We propose an enrollment rule that prescribes the maximum number of patients that may be enrolled at any given point in the trial. Key words: Delayed outcome, Phase II oncology trial, Pocock boundary, Stopping rule, Enrollment rule.

Keywords

Delayed outcome; Phase II oncology trial; Pocock boundary; Stopping rule; Enrollment rule

1. Introduction

Many oncology Phase II trials implement a stopping rule for toxicity. This is because the toxicity profile of a drug or a drug combination used in a Phase II trial might not be well understood by the time the trial commences. A number of stopping rules for toxicity have been proposed for use in a single-arm trial. Ivanova, Qaqish and Schell (2005) argued for the use of the Pocock type (Pocock, 1977) stopping boundary; Geller et al. (2005) proposed a Bayesian stopping rule. Both argued that a continuous stopping rule, a rule that can be applied after any number of patients have completed follow-up for toxicity, provides the best protection against observing an excessive number of toxicities and therefore is preferable to two or three-stage stopping rules. A continuous stopping rule is a multistage rule where the number of stages is the same as the number of patients in the trial. The continuous Pocock boundary is routinely used in Phase II oncology trials conducted by the Lineberger Comprehensive Cancer Center (LCCC).

In some oncology trials the follow-up for toxicity is long compared to the accrual rate. This was the case in a Phase II study of a novel B-Raf inhibitor administered together with a

Address correspondence to: Anastasia Ivanova, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, U.S.A., aivanova@bios.unc.edu.

monoclonal antibody in patients with active melanoma brain metastases. The observation period for dose limiting toxicity (DLT) was 12 weeks. The Pocock boundary was in place to monitor the DLT rate. To use the Pocock boundary one needs to have full follow-up data on all already enrolled patients. If investigators in this study wait for all patients to be fully followed before the next patient is enrolled, the length of the trial with 30 patients will be over 7 years. Is it possible to make the trial shorter without increasing the number of patients potentially exposed to unsafe treatment? The problem can be solved 1) by developing a stopping boundary that utilizes all data available including data from patients still in follow-up, and 2) by guiding the accrual rate to the trial so that future enrollment decision is determined by the observed toxicity pattern in the previous enrolled patients. The problem of conducting studies with long follow-up arises in Phase I oncology trials as well (Cheung and Chappell, 2000, Bekele et al., 2008). In order to utilize the partial data in a Phase I dose-finding trial, Cheung and Chappell (2000) used an assumption about the distribution of time to toxicity in $(0, t^*)$ given toxicity occurs in $(0, t^*)$, where t^* is the length of follow-up for toxicity. Bekele et al. (2008) suggested halting enrollment to a Phase I trial when there is not enough data to select a safe dose for the next patient. We extend these ideas to a single arm Phase II oncology trial. Our first contribution is the method of deciding whether to stop the trial or not based on all available data, including data from patients still in follow-up. The main challenge here is to maintain the pre-specified probability of stopping the trial under various true toxicity rates, in particular, maintaining the desired type I error rate and power, when partial data are used. Our second contribution is an enrollment rule that prescribes, based on current data, when to enroll a patient, how many to enroll, and when to wait. In this paper, in Section 2 we develop a continuous stopping rule for toxicity based on partial data. Enrollment strategy is described in Section 3, and a simulation study is presented in Section 4. We give a real trial example in Section 5 and present conclusions in Section 6.

2. Stopping for toxicity based on partial data

In a Phase II study each patient is followed for toxicity for a fixed period time of t^*. Let n be the number of patients enrolled in the study so far. Let U_i be the random variable denoting the time to toxicity for the ith patient and let $\theta = P(U_i \le t^*)$ be the probability of toxicity. Denote $Y_{i,n}$ the indicator that the ith patient has experienced toxicity by the time just prior to the entry of the $(n+1)$th patient, $i = 1,\ldots,n$, then $X_n = \sum_{i=1}^{n} Y_{i,n}$ is the random variable denoting the total number of toxicities observed at that time.

In a trial where all patients are fully followed for toxicity when the next patient is assigned, $Y_{i,n}$ does not depend on n and follows a Bernoulli(θ) distribution and $X_n \sim$ binomial(n, θ). Ivanova et al. (2005) considered such trials and argued that the Pocock boundary is the most suitable boundary for monitoring toxicity in a Phase II oncology trial as it allows stopping early with high probability and therefore reduces the expected number of toxicities. Let K be the sample size planned for a Phase II study and let θ_0 be the acceptable probability of toxicity. The Pocock boundary can be defined through a point-wise probability α, such that the trial is stopped if, at each interim analysis, the null hypothesis $\theta = \theta_0$ is rejected at level α in favor of the one-sided alternative $\theta > \theta_0$. The value of α is chosen so that the overall probability of stopping the trial, φ, is equal to a specified value, usually $\varphi = 0.05$, if the

toxicity rate is θ_0. In another words, the type I error is controlled at level φ. We refer to the boundary that allows stopping the study at any point as a continuous boundary, because the trial can stop after any number of patients are accrued and their results observed. Let the constants b_k, $k = 1,\ldots,K$, be the smallest integer such that $P[X_k \geq b_k] \leq \alpha$, then such a boundary can be described through (b_1, b_2, …, b_K). If the number of toxicities in the first k patients is equal to or higher than b_k, the trial is stopped. Another way of implementing this boundary is to compute a one-sided p-value to test the null hypothesis $\theta = \theta_0$ versus one-sided alternative $\theta > \theta_0$ after each patient's outcome is observed. The trial is stopped if the p-value is less than α. In fact, it is sufficient to apply the boundary only when a patient experiences toxicity. This is because, if the trial has not been stopped before, it can only be stopped if more toxicities are observed.

We now explain how to use data from partially followed patients to implement a sequential boundary. We will compute the p-value to test the null hypothesis $\theta = \theta_0$ versus the one-sided alternative $\theta > \theta_0$ based on all information available using assumption on the distribution of time to toxicity in (0, t^*).

Now consider the case when not all n patients are fully followed for toxicity at the time just prior to the entry of the (n+1)th patient. Let $t_{i,n}$ be time elapsed from the start of treatment for the ith patient at the time just prior to the entry of the (n+1)th patient. Following Cheung and Chappell (2000), for $t_{i,n} < t^*$, we have

$$P(U_i \leq t_{i,n}) = P(U_i \leq t_{i,n} | U_i \leq t^*) P(U_i \leq t^*) = P(U_i < t_{i,n} | U_i \leq t^*)\theta = \frac{t_{i,n}}{t^*}\theta = w_{i,n}\theta.$$

In other words, a weight $w_{i,n} = P(U_i < t_{i,n} \mid U_i \leq t^*)$ is assigned to the ith patient and the probability that the ith patient experiences toxicity when treated for a length of $t_{i,n}$ is $w_{i,n}\theta$. This is equivalent to assuming that the time to toxicity given that toxicity occurs in (0, t^*) follows a uniform distribution on the interval (0, t^*). This and other weighting options were described in Cheung and Chappell (2000) and Yin (2012). The weight $w_{i,n}$ is set to 1 for patients who have already experienced toxicity and patients who have completed follow-up. At the time just prior to the entry of the (n+1)th patient, the number of patients who completed the trial without toxicity is $S_n = \sum_{i=1}^{n} I(w_{i,n} = 1) - X_n$ and the number of patients still under follow-up is $R_n = \sum_{i=1}^{n} I(0 < w_{i,n} < 1)$, $X_n + S_n + R_n = n$. Let x, s and r denote the observed X_n, S_n and R_n respectively. The one-sided p-value for testing the null hypothesis $\theta = \theta_0$ versus $\theta > \theta_0$ is the probability of $X_n \geq x$.

For example, if $n = 3$, $x = 2$, $s = 1$ and $r = 0$, then all weights are equal to 1 and the p-value is calculated as $P[X_3 \geq x] = P[X_3 \geq 2] = 0.104$, where X_3 is a binomial random variable with parameters $n = 3$ and $\theta_0 = 0.2$, $X_3 \sim$ binomial(3,0.2). As another example, if the counts right before enrolling the fourth patient are $n = 3$, $x = 2$, $s = 0$ and $r = 1$ with the first two patients fully followed and time $t_{3,3} = t^*/2$ and hence $w_{3,3} = 1/2$ for the patient still in follow-up, then $X_3 = Y_{1,3} + Y_{2,3} + Y_{3,3}$, where $Y_{i,3} \sim$ Bernoulli(θ_0) for $i = 1,2$ and $Y_{3,3} \sim$ Bernoulli($\theta_0 / 2$). Therefore

$$P[X_3 \geq 2] = P[Y_{1,3} + Y_{2,3} \geq 1, Y_{3,3} = 1] + P[Y_{1,3} = Y_{2,3} = 1, Y_{3,3} = 0] = 0.072.$$

We illustrate the ability of this rule to stop the trial via simulations. Consider an example of a Phase II trial with $K = 30$. To yield the overall probability of stopping of 0.05 when $\theta_0 = 0.2$, we need to use $\alpha = 0.0164$ at each step. The continuous Pocock boundary (Pocock, 1977) for this trial is shown in Table 1. In our simulations study $t^* = 12$ weeks and a new patient is enrolled every week. In reality, time to toxicity might not follow uniform distribution on $(0, t^*)$. Time to toxicity can be more generally modelled using Weibull distribution with corresponding survival function $S(t) = e^{-\lambda t^\gamma}$. When the shape parameter γ is 1, it is the same as exponential distribution. When $\gamma > 1$, toxicity is more likely to occur at the end of the interval. Table 2 contains results for the proposed stopping rule when time to toxicity is distributed according to uniform, exponential ($\gamma = 1$) and Weibull distribution with $\gamma = 2$. For comparison, we also present data for a trial with instantaneous toxicity outcome, $t^* = 0$, and for a trial with $t^* = 12$ weeks where only data from patients who were fully followed (i.e., full follow up time t^* passed from the initiation of treatment for all patients irrespective of their outcome) is used in a stopping rule. As seen from Table 2, our rule allows to stop the trial earlier and to observe less toxicity on average, especially when toxicity rate is high. The rule based on partial data yields slightly lower type I error rate (i.e. stops less frequently when the true toxicity rate is low) and power (i.e., stops less frequently when the true toxicity rate is high), however the decrease is not substantial. Even though the proposed stopping was based on the assumption that time to toxicity follows uniform distribution, it performs well when the true distribution of time to toxicity is exponential or Weibull. The method is robust towards the distribution of time to toxicity because eventually each patient completes their follow-up for toxicity and their toxicity outcome becomes known. The imputation based on uniform distribution works fine for those patients who eventually complete the follow-up and slightly loses accuracy for those patients do not complete the follow-up. If needed, one can use a more flexible model for time to toxicity (Cheung and Chappell, 2000).

3. Enrollment rule to prevent excessive number of toxicities

If many patients are enrolled at once, the stopping rule described in the previous section will not prevent assigning too many patients to a regimen that may not be safe. Often, many patients are enrolled at the very beginning of the trial which might lead to excessive toxicities. An enrollment rule informs investigators about how many patients may be enrolled at the beginning of the trial and guides further accrual based on the information about toxicity in the trial.

Consider the boundary in Table 1. Initially we may enroll 3 patients as it is not possible to stop the trial before 3 patients complete follow-up. If none of these patients experience toxicity in $(0, t^*)$, one may enroll as many as 5 more patients, since there is a possibility to cross the boundary by observing 5 toxicities out of 8 patients, and it is not possible to cross the boundary if less than 5 additional patients are enrolled. More formally, the trial can enroll m new patients such that $r + x + m \leq b_{n+m}$, $r + x + m - 1 < b_{n+m-1}$ and $n + m \leq K$. The

assumption here is the worst case scenario, i.e., the toxicity rate is $\theta = 1$ and therefore every patient in the follow-up will experience toxicity. This assumption was referred to as the conservative plan in Schmegner and Baron (2004) who considered it in the context of sequential planning of experiments. This is the most conservative enrollment rule and the number of toxicities we observe will be very similar compared to the trial with instantaneous toxicity response. However, this rule can lead to a rather long trial if t^* is long. We propose a way to relax this rule.

Let M be the design parameter fixed in advance. One can think of M as the maximum number of extra toxicities we are willing to allow in order to make the trial shorter. The maximum number of new patients to enroll, m, is determined by $r + x + m \leq b_{n+m} + M$, $r + x + m - 1 < b_{n+m-1} + M$ and $n + m \leq K$. That is, at any time the maximum number of patients experiencing toxicity cannot exceed the number allowed by the Pocock boundary plus M. As before we assume the worst case scenario that all patients will experience toxicity and allow M extra toxicities beyond what is allowed by the Pocock boundary. When $M = 0$, the rule is equivalent to the conservative enrollment plan from Schmegner and Baron (2004). The maximum number of patients to enroll in the trial initially is $b^* + M$, where b^* is the minimum number k such that $k \geq b_k$. In the example in Table 1, we can enroll at most $3 + M$ patients initially. If M is as large as $M \geq K - b^*$, all patients can be enrolled in the beginning of the study. This strategy has a clear interpretation as allowing at most M additional toxicities over what is allowed by the stopping rule. We will refer to this strategy as "+M enrollment rule" in the remainder of the paper.

As mentioned earlier, using just the stopping or just the enrollment rule will not prevent the trial from possibly seeing excessive toxicity. The algorithm below describes how to apply both the stopping rule from Section 2 and the +M enrollment rule described in this Section in a clinical trial.

- **i.** Initial enrollment is $b^* + M$. For example, in Table 1, $b^* = 3$.
- **ii.** When toxicity is observed, calculate the p-value as described in Section 3.3. If p-value is less than α, stop the trial.
- **iii.** When there is a toxicity or a patient reached the end of follow-up t^* without toxicity, if $x + r \geq b_{x+s+r} + M$, no new patients may be enrolled. If $x + r < b_{x+s+r} + M$ and the enrollment limit has not been reached, find the smallest integer m that satisfies $m \geq b_{x+s+r+m} + M - (x + r)$. If $m > x + r + s$, then enroll $K - (x + r + s)$ patients, otherwise, enroll m patients.

4. Simulation results and discussion of design parameters

In this section we present a simulation study investigating the performance of the +M enrollment rule for various values of M in conjunction with the stopping rule described in Section 2. We used the example from Section 2 with $K = 30$ and $\theta_0 = 0.2$. Figures 1–3 show the expected number of toxicities, expected number of patients enrolled and expected length of trial (in units of t^*) for some values of M across the range of true toxicity rate. When $M = 0$ the probability of stopping the trial is almost the same as φ. As M increases, assuming that patients are always available to enroll in the trial, the probability of stopping the trial

decreases slightly. At the same time, the expected number of toxicities is increasing, mostly because many more patients are enrolled before the trial is stopped and not because the probability of stopping gets slightly lower compared to when $M = 0$. On the other hand, as M increases the trial gets shorter. Note that when $M = 27$ all patients may be enrolled at the beginning of the trial.

To choose an appropriate M for a trial, we notice that for a given M, the expected number of toxicities rises as the true toxicity rate increases. If toxicity rate $\theta = 1$, the increase in expected toxicity compared to a sequential trial with instantaneous response is M. For $\theta < 1$ the expected increase will never exceed θM. Furthermore, because a stopping rule is in place, the expected increase is smaller than θM. One can estimate the expected increase in toxicity for each stopping and enrollment rule combination and t^* by simulations. Consider the example in Table 1 and assume that the true toxicity rate cannot be higher than $\theta = 0.6$. Simulations show that if we chose $M = 5$ in the $+M$ enrollment rule, we will see at most one extra expected toxicity when $\theta \leq 0.6$. Therefore $M = 5$ is a good choice of parameter value in the $+M$ enrollment rule if we are willing to allow at most one extra toxicity.

5. Real trial example

The proposed methodology is used in ongoing LCCC pharmacokinetic study of patients with high risk myelodysplasia and acute leukemia. Dose limiting toxicity outcomes (yes or no) were defined as non-relapse mortality or one of the following toxicities observed during the first 8 weeks from the start of treatment: grade 3 non-hematologic toxicity lasting greater than 7 days, grade 4 non-hematologic toxicity with the exception of drug-related fever, or grade 3/4 hematologic toxicity lasting greater than 42 days. The total number of patients in the study was 22. The probability of stopping the trial was set to 0.05 when the true dose limiting toxicity rate is equal to the tolerable rate of 0.2. The investigators preferred the most conservative approach to enrollment and therefore M was set to 0 in the $+M$ enrollment rule. This enrollment rule allowed enrolling 3 patients in the beginning of the study. There were no dose limiting toxicities in the first 3 patients, therefore 5 more patients could be enrolled at once. After enrolling the first three patients the accrual has slowed down. The study is ongoing.

6. Conclusions

We propose a frequentist sequential stopping rule for toxicity that utilizes all available data in the trial. To control the number of toxicities in the study, we recommend using the stopping rule with an enrollment strategy. The parameter to use in the enrollment strategy can be chosen based on the maximum number of extra toxicities or to yield a desired trade-off between the length of the trial and an increase in expected number of toxicities. The continuous sequential boundary (Ivanova et al., 2005) can be generated by using software available at http://cancer.unc.edu/biostatistics/program/ivanova/.

Acknowledgement

This work was supported in part by NIH grant RO1 CA120082-01A1. The authors thank Dr. Matt Foster for the permission to use the LCCC study in patients with high risk myelodysplasia as an example.

The authors thank an associate editor and anonymous reviewers for their helpful comments.

References

1. Bekele BN, Ji Y, Shen Y, Thall PF. Monitoring late-onset toxicities in phase I trials using predicted risks. Biostatistics. 2008; 9:442–457. [PubMed: 18084008]
2. Cheung YK, Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. Biometrics. 2000; 56:1177–1182. [PubMed: 11129476]
3. Geller, NL.; Follmann, DF.; Leifer, ES.; Carter, SL. Design of early trials in peripheral blood stem cell transplantation: a hybrid frequentist-Bayesian approach. In: Geller, NL., editor. Advances in Clinical Trial Biostatistics. New York and Basel: Marcel Dekker; 2005. p. 40-52.
4. Ivanova A, Qaqish BF, Schell MJ. Continuous toxicity monitoring in phase I trials in oncology. Biometrics. 2005; 35:540–545. [PubMed: 16011702]
5. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika. 1977; 64:191–199.
6. Schmegner C, Baron M. Principles of optimal sequential planning. Sequential Analysis. 2004; 23:11–32.
7. Yin, G. Clinical Trial Design: Bayesian and Frequentist Adaptive Methods. New Jersey: John Wiley & Sons; 2012.

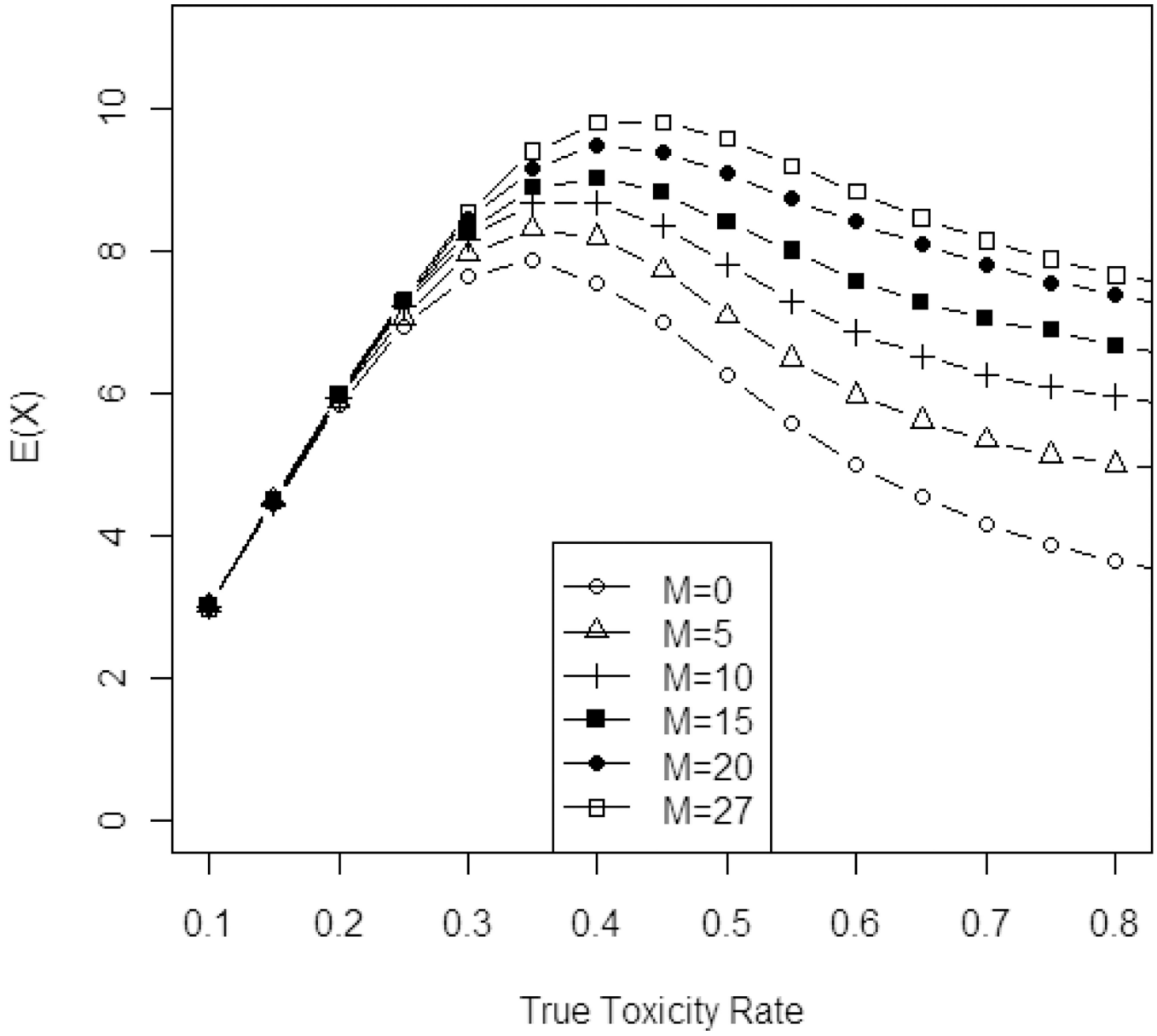

Figure 1.
Expected number of toxicities plotted versus true toxicity rate for different values of M in a trial with $K = 30$, $\theta_0 = 0.2$ and $\varphi = 0.05$.

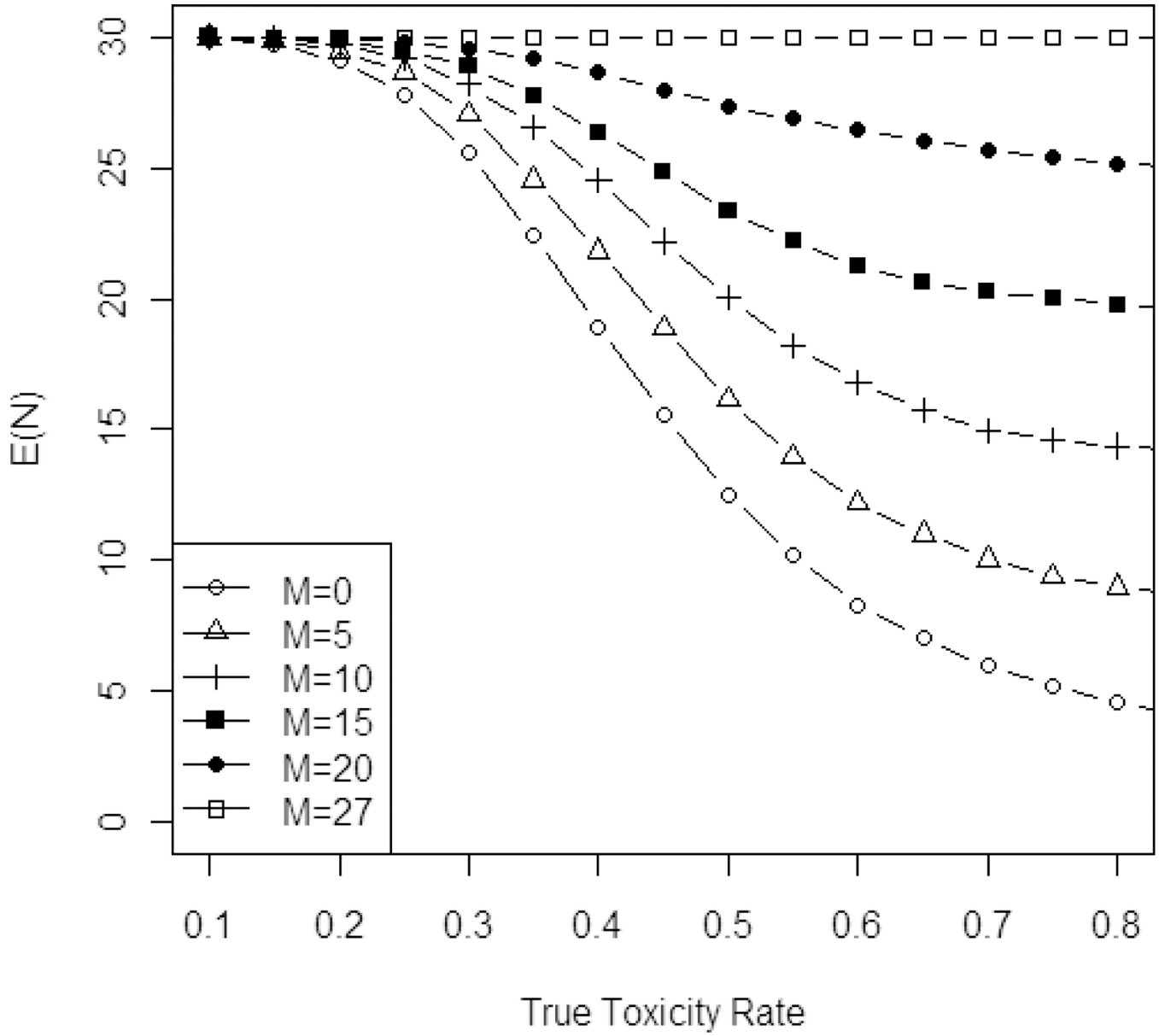

Figure 2.
Expected number of patients enrolled in the trial versus true toxicity rate for different values of M in a trial with $K = 30$, $\theta_0 = 0.2$ and $\varphi = 0.05$.

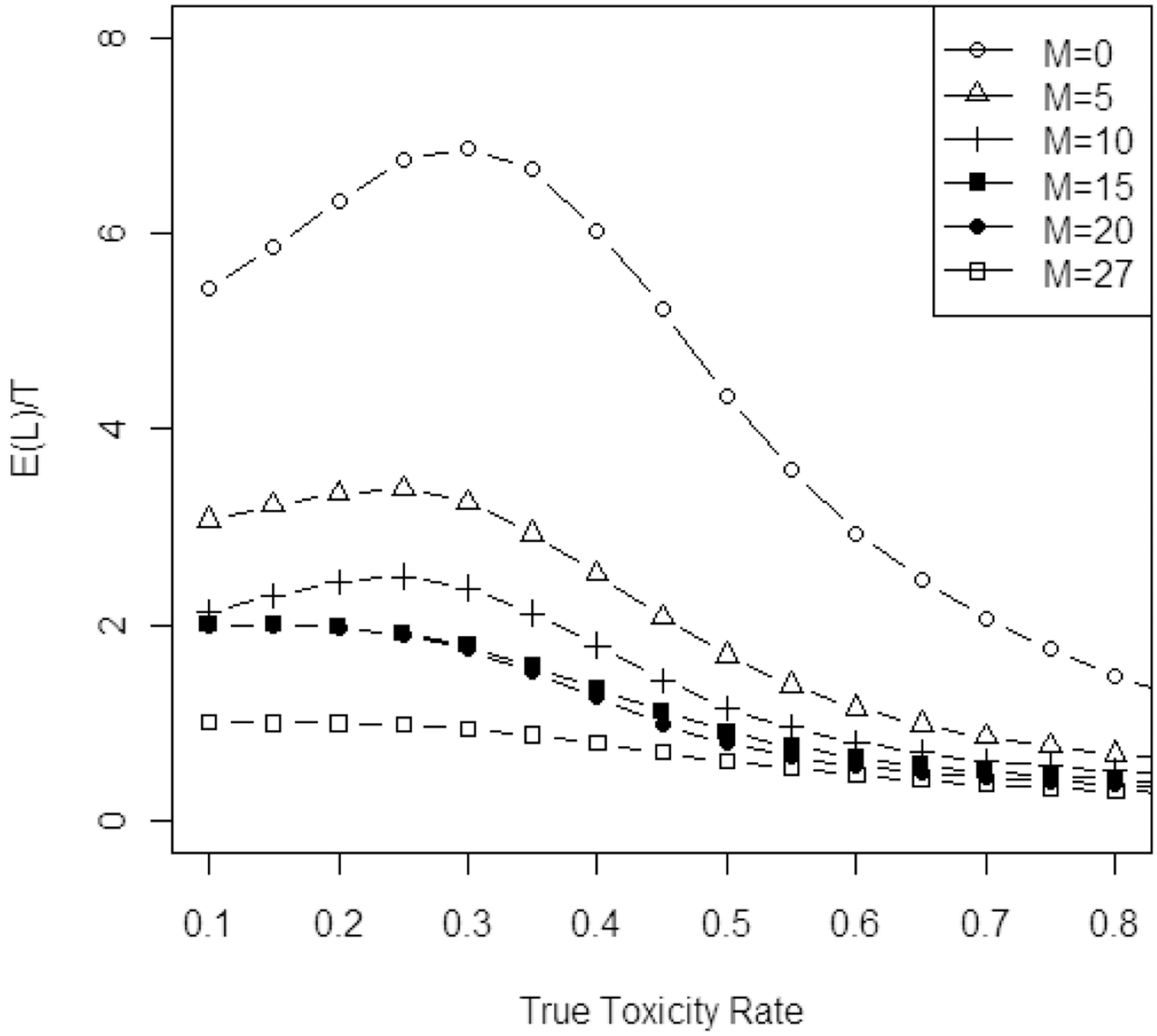

Figure 3.
Expected length of study versus true toxicity rates for several values of M in a trial with K = 30, $\theta_0 = 0.2$ and $\varphi = 0.05$.

Table 1

The Pocock stopping boundary $\{b_k\}$ for $K = 30$, $\theta_0 = 0.2$ and $\varphi = 0.05$ yielding $\alpha = 0.0164$,

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
b_k	--	--	3	4	4	4	5	5	6	6	6	7	7	7	8
k	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
b_k	8	8	8	9	9	9	10	10	10	11	11	11	11	12	12

J Biopharm Stat. Author manuscript; available in PMC 2016 September 18.

Table 2

Comparing the new stopping rule that uses all available data when time to toxicity is uniformly distributed in $(0,t^*)$ (Uniform), exponentially distributed (Expon), has Weibull distribution with shape parameter 2 (Weib(2)) with the rule that uses fully followed patients only (Full data) for a trial that $K = 30$, $\theta_0 = 0.2$, and $t^* = 12$ weeks with one patient being enrolled every week. For comparison we show results for a trial with instantaneous response ($t^* = 0$). We display the probability of stopping the trial, the expected number of patients enrolled before the trial is stopped, E(N), and the expected number of toxicities, E(X).

True toxicity		**0.2**	**0.4**	**0.6**	**0.9**
Probability of stopping	$T = 0$	0.05	0.70	1	1
	Uniform	0.03	0.68	1	1
	Expon	0.03	0.69	1	1
	Weib(2)	0.02	0.64	0.99	1
	Full data	0.03	0.61	0.99	1
E(N)	$T = 0$	29.1	19.0	8.3	4.6
	Uniform	29.6	23.8	15.2	10.0
	Expon	29.6	23.0	13.6	7.1
	Weib(2)	29.9	25.9	17.3	10.5
	Full data	29.8	27.1	21.0	15.8
E(X)	$T = 0$	5.8	7.6	5.0	3.6
	Uniform	5.9	8.4	6.2	4.6
	Expon	5.9	8.1	5.7	3.9
	Weib(2)	6.0	8.3	8.0	4.7
	Full data	5.9	9.8	9.4	9.3