Statistical design of Phase I clinical trials

by

Weijia Zhang

A thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements for the degree of

Master of Science

Department of Statistics
The University of Manitoba
Winnipeg

To My Parents

Abstract

There are four phases of a successful clinical trial. Phase I determines the maximum tolerated dose (MTD) of the drug in human beings. Phase II trial examines the new drug's short term efficacy and is the proof-of-concept trial for Phase III. Phase III is a large and long term trial before applying for marketing the drug. It concerns the long-term efficacy and safety of the drug. If the new drug is approved for marketing, Phase VI is the post-marketing surveillance trial. The goal is to monitor potential side effects of the new drug.

My MSc research is focused on the design of Phase I clinical trials. There are many designs available in the literature for Phase I clinical trials, including both nonparametric and parametric designs. The most famous parametric design is the continual reassessment method (CRM). A parametric model is assumed and has unknown parameters. These unknown parameters follow a prior distribution under the Bayesian approach. The observations of patients treated are either toxic or nontoxic. Observations of patients' toxicities are used to update the prior distribution into a posterior distribution. Allocation of the dose to the next patient is adaptive and based on the estimated toxicity probability that is near the desired probability. The objective is to identify the maximum tolerated dose to be used in Phase II clinical trials.

Three parametric models are normally used with the continual reassessment method, namely the logistic, power and hyperbolic models. They use respectively the logistic, power and hyperbolic functions. These functions are used to define the dose probabilities at different dose levels, but they are not so flexible.

In this thesis, we use the continual reassessment method with a new class of parametric functions. This class is formed with the cumulative distribution function of the normal distribution. The major advantage is that we can choose different values of the mean and variance of the normal distribution so we can model different shapes of dose toxicity probability relationship. So this new class of parametric designs is very flexible. We conduct simulation studies and compare our new design with the existing parametric designs. We have found our design performs better by choosing the appropriate values of the mean and variance. The use of the cumulative distribution function was motivated by the Advanced Statistical Theory course taught by Dr. Liqun Wang.

When the toxicity is not immediately observable but delayed, we extend the continual reassessment method with the cumulative distribution function to late-onset toxicities. We assume the time to toxicity follows a geometric distribution. We conduct simulation studies and compare our new design with the existing parametric designs. We have found our new design with late-onset toxicities performs better by choosing the appropriate values of the mean and variance. The use of the geometrically distributed late-onset toxicity was motivated from taking the Lifetime Data Analysis course taught by Dr. Po Yang.

Acknowledgements

First of all, I sincerely thank my thesis advisor Dr. Po Yang for her patient guidance and constant support during this research. I thank members of my Thesis Committee, Dr. Saumen Mandal and Dr. Jun Cai, for reading my thesis and offering constructive criticisms.

I thank Dr. Liqun Wang for teaching us the Advanced Statistical Theory. The main idea of my new design was motivated by the statistical theory from this course when I realized that the cumulative distribution function may be used with the continual reassessment method. I thank Dr. Po Yang for teaching us the Lifetime Data Analysis course. The idea of extending my new design to late-onset toxicity came from taking this course. I thank Dr. Saman Muthukumarana and Dr. Alex Leblanc for teaching us R programming. From these two courses, I learned R programming skills which have been very useful for my simulations. I thank Mr. Dave Gabrielson for teaching us Latex programming. Finally, I thank Dr. Xikui Wang for reading and correcting my thesis, and for his constructive criticisms.

I gratefully acknowledge financial supports from the Faculty of Graduate Studies, the Faculty of Science, the Department of Statistics, and Dr. Yang's research grant.

Finally I thank my parents for their love and support, and all members of the Department of Statistics and many friends for their support.

Contents

Chapter 1

Introduction

In this Chapter, we review some basic concepts about the statistical design of Phase I clinical trials.

1.1 Introduction

We all want to live long and healthy, but in reality many people are threatened by health problems, such as heart disease, cancer, diabetes and so on. The goal of medical research is to find effective treatments to cure diseases and improve the quality of life of human beings.

Statistics show the global life expectancy was 46 years in 1950 but increased to 75 years in 2015 (UnitedNations, Accessed 2016-06-28). Advance in medicine has played an important role in prolonging life besides our better living conditions. Clinical trials are a necessary part of modern medical research, and are experimental designs to evaluate the reliability and effectiveness of new medical

interventions in human beings.

1.2 Scientific inquiry and study methods

There are three progressive classes of knowledge in empirical science: observed phenomena among events; association among phenomena; and at the most advanced level, causation between phenomena (Rosenberger and Lachin, 2002). In medical research, it is important to identify the causal relationship between disease and risk factors, such as genetic variation, environmental and living conditions and some bad habits. For example, why do smokers have more chance to expose lung cancer than non-smokers, and why does not every smoker have lung cancer? The possible answer is that smoking is one of the factors causing cancer, but not the only factor. The task of scientists is to find all possible factors resulting in diseases and to prevent or cure diseases. How can scientists find these causations and prevent and treat diseases? One study method is to get information and data from alternative treatments. Clinical trials are such experimental designs to statistically test whether new drugs are significantly more effective than the old ones.

1.3 Preclinical and clinical research

In preclinical and early phase clinical trials, a small number of patients are treated with safe doses to investigate the pharmacokinetics and pharmacodynamics of

the new drug (Hedaya, 2007), that is, we need to study what the drug works for the body and how the body responses to the drug. Compared to later phases of clinical trials, the preclinical trials have no therapeutic intent (Kummar et al., 2008)

After the preclinical trials, we move on to the first-in-human phase I clinical trials. The goal of phase I clinical trials is to identify the maximum tolerated dose (MTD) in human beings for a new drug. Any drug has two sides. The drug has benefits of treating disease and the risk of toxicity. A Phase I clinical trial is normally conducted on 15 to 30 patients to determine the safe dose (MDAnderson, Accessed 2016-06-28), that is the maximum dose given to patients who can accept the toxicity level. We usually assume the toxicity monotonically increases by the dose level, so the trial starts from a low and safe dose, and escalates or de-escalates the dose level depending on whether the treated patients experience toxicity or not. The determined MTD will be investigated further in Phase II and Phase III trials for its efficacy. Any over-dose or under-dose of MTD will mislead the drug research.

In phase II trials, we study the new drug's short-term efficacy. This is normally called the proof-of-concept trial for Phase III. Only a drug showing potential effect enters a phase III trial, otherwise the study is terminated. For example, if a tumour shrinks by 50% or more, the longest diameter of target lesion decreases by 30% or the whole target lesion disappears in comparison with

baseline treatment, the treatment is regarded to be potentially successful. It takes a relatively short time to finish a phase II trial and start smoothly a phase III trial.

Phase III trial is the most important and the last stage before applying for marketing the drug. It concerns the long-term efficacy and safety of the drug. It requires a very large number of patients (normally ranging from hundreds to thousands of patients) and a control treatment. Patients are randomized to receive one of alternative treatments. Only if the new drug shows significant benefit over the other treatments, it may be approved to enter the market.

After a successful Phase III trial, if the new drug is approved by FDA (Food and Drug Administration in the United States), a post-marketing surveillance Phase VI trial starts. The purpose is to monitor potential side effects. If side effects are severe, the drug may be required to withdraw from the market.

1.4 Motivation

Clinical trials are important for developing new medical interventions including drugs, and have remained to be an important part of mainstream clinical research. Drug development has a vital in improving both quality and quantity of our lives. However the process of developing new drugs is long and expensive, and many important statistical factors and ethical issues have to be taken into consideration. Clinical trials are designed experiments on human subjects and must be safe

and ethical. Monitoring toxicity and side effects are important issues. Phase I clinical trials are the building blocks for late phase clinical trials, and it is important to derive the most reliable, most safe and most effective dose for the best potential benefit of the drug. Under certain assumptions, this best dose is the maximum tolerated dose (MTD) that is to be used in Phase II and Phase III clinical trials. Therefore among the four phases of clinical trials, Phase I is fundamentally important because both over-estimation and under-estimation of the maximum tolerated dose will lead to over failure of the clinical research.

Although there are many different approaches to the design of Phase I clinical trials, each of the designs can be improved. The non-parametric designs are easy to implement, however not necessarily optimal. In any case, the non-parametric approach assumes no particular form of the dose toxicity probability relationship. Moreover the decision of escalation or de-escalation of the next dose selection depends on the current dose and response only, and ignores all other previous doses selections and their associated responses. On the other hand, the parametric designs assume particular forms of the dose toxicity probability relationship, however more tedious to implement. For the parametric approach, the dose toxicity probability relationship is described by a distribution function that defines the toxicity probability at each dose level. This function describes the overall relationship between toxicity probabilities and the doses, and contains unknown parameters. These unknown parameters can be estimated using either the Bayesian approach or the frequentist method. Since making inference about

the whole curve describes the toxicity probability, the decision of dose selection for the next patients is determined by all past doses and their associated responses. These parametric models are incorporated with the continual reassessment method (CRM) to assign doses to patients and estimate the maximum tolerated dose (MTD).

In the parametric approach, the particular toxicity probability function is very important because it defines the toxicity probability at each dose. Currently in the literature, there are mainly three types of parametric functions used with CRM: the logistic model, the power model, and the hyperbolic tangent model. All these functions share two common characteristics: they are increasing and change concavity from concaving up to concaving down. When I took the graduate course on Advanced Statistical Inference from Dr. Liqun Wang, the cumulative distribution function $F(x)$ is used often, particularly for the standard normal distribution function. So if we replace the function $exp(x)$ with the cumulative distribution function $\Phi(x)$ for the standard normal distribution and use the function $p(x) = \frac{2\Phi(x)}{1+\Phi(x)}$ in combination with an unknown parameter, we may also apply CRM. Indeed, this function is increasing and changes concavity, and has values between 0 and 1. I then thought about generalizing further by considering the normal distribution with different variances. This became the motivation for writing Chapter 3 of my thesis.

In reality, if a patient is toxic, we may not be able to observe toxicity

immediate after the treatment. That is, the toxicity of treating patient may be late onset. I still treated two possibilities. The patient may be non-toxic at all, or the patient may be toxic. If the patient is toxic, we assume that the observation is late-onset. That is, the time to observe the toxic response after treatment is delayed and maybe censored when the next patient is treated. Because we model patients who are treated one after another in discrete time, this requires me to think of a discrete distribution to model waiting time. When I took the graduate course on Survival Analysis from Dr. Po Yang, I learned some survival distributions. So I thought the geometric time distribution can be a good fit for describing the time to toxicity if the patient is toxic after treatment at a particular dose. This motivated me to work on Chapter 4.

I have always been interested in biostatistics because I believe it is not only very useful but also practical to find good jobs. One of the major fields of research in biostatistics is the design and analysis of clinical trials, therefore I wanted to work on Phase I clinical trials for my MSc research because it is the most fundamental phase in clinical trials. After my above described motivations, I believed that the new toxicity probability function might perform good and might be even better than the ones in the literature. The new function is also flexible because we can change the variance of the normal distribution, so we have a class of toxicity probability functions. It is impossible to show this theoretically, so I decided to do lots of simulations to see if our new design is better. After extensive simulations, we have observed that our new toxicity function works better than

the other functions in the literature. This new function can be significant because if we use it in practice, we will be able to treat more patients in the trial better and still determine the MTD from the clinical study. That is, we are more ethical to the patients in the trial and at the same time do a better preparation for the patients in the late phases of the clinical research.

1.5 Summary and structure of the thesis

The goal of the thesis is to introduce a new parametric design for Phase I clinical trials and assess its performance. We consider both immediate and late-onset toxicities. In this chapter, we have reviewed some background materials for Phase I clinical trials.

In Chapter 2, we review both nonparametric and parametric designs of Phase I clinical trials.

In Chapter 3, we introduce a new model for the relationship between dose and its toxicity probability. We apply this new model to the parametric design using the continual reassessment method. The performance of this new method is compared with existing parametric methods. Extensive simulation results are derived to compare the new method with existing methods.

In Chapter 4, we assume late-onset toxicity and use the geometric distribution to model the delayed time to toxicity. Then the continual reassessment method is modified to incorporate late-onset toxicity. Extensive simulation results are

derived to compare the new method with existing methods.

Chapter 5 concludes the thesis by summarizing the results and discussing future research problems.

Chapter 2

Design of phase I clinical trials

2.1 Introduction

Phase I trial is the first stage of a drug test in human beings before the approval of the drug for marketing. It is very crucial in the entire process of drug development because we need to recommend the maximum tolerated dose (MTD) to Phase II and phase III clinical trials. If the MTD is over-estimated, patients may be exposed a very risky treatment in the later trials. If the MTD is under-estimated, the treatment may not be effective and we concern more about the toxicity than the efficacy of the new drug. Trade-off between efficacy and toxicity of the new drug is our goal in all the steps of clinical trials. Phase I trial is only focused on the maximum toxicity dose that patients are not able to tolerate any more. Generally speaking, from 15 to 30 patients are involved in the Phase I trial (MDAnderson, Accessed 2016-06-28), which starts from a very low and safe dose specified from the animal test (usually a small fraction of the MTD in animals). Patients are

treated in cohorts, say of size three in each cohort. The first cohort is treated at the starting dose. If a large number of patients in this cohort experience toxicity, the next cohort will be treated as a lower dose. Otherwise the next cohort is given a higher dose. We expect to observe toxicity immediately for decision. The situation that the toxicity is not observable from the previous cohort until the next cohort greatly impacts on the decision-making of dose selection. We face survived data when we need to determine whether or not to escalate or de-escalate the dose level for the next cohort. This topic is discussed in Chapter 4 of my thesis.

As we know, the goal of a Phase I clinical trial is to identify the MTD for later phase studies. Some important issues of this trial are taken into account: ethical concerns, the starting dose, the speed of dose escalation, the sample size of the patients, and the target toxicity probability. These issues have direct impact on the efficiency of Phase I trial design and the accuracy of the identified MTD level.

Dose-finding methods in Phase I trial are classified into two classes: one class consists of nonparametric methods, including 3+3 designs and A+B designs (Storer, 1989), the accelerated titration design (ATD) (Simon et al., 1997) and biased coin dose-finding method (BCD) (Durham et al., 1997). The other is the class of parametric methods which includes continual reassessment method (CRM) (O'Quigley et al., 1990).

2.2 Nonparametric methods

2.2.1 3+3 Methods and extension

In Phase I trial study, we are responsible for both safety and efficacy of the treatment. So the dose cannot be overly toxic and overly low to be effective. That means escalation of dose cannot be too fast or too slow.

The 3+3 design is a standard and classic method that identifies the MTD with a targeted toxicity probability. Storer (1989) shows with such design, the toxicity probability will be less than 33%. In a real clinical trial, the 3+3 design is widely utilized because of its easy implementation. We assume a predefined dose level d_i is increasing in the trial study and the corresponding toxicity probability p_i is non-decreasing. The cohort size is 3, and the first 3 patients are treated at the lowest starting dose level. The process of 3+3 design is described as follows:

(1) Suppose three patients are treated at a current dose level i, and we need to evaluate for toxicity.

(2) If no patient is toxic, escalate to the next higher dose level $i+1$, and go back to step (1).

(3) If one patient is toxic, then three more patients are treated at this same dose level i and we observe dose limiting toxicity (DLT) of the total of six patients. If only one patient of the group of six patients experiences toxicity, we move to the next higher dose level $i+1$. If two patients of these six

patients are toxic, the trial is terminated and the next lower dose level $i-1$ is considered as the MDT. If more than two patients of these six patients are toxic, then the current dose level i is greater than the MTD, and another three patients will enter the treatment and will be treated at the next lower dose level $i-1$.

(4) If more than one patient is toxic the lowest dose level, the trial will be terminated and is said to be an inconclusive trial.

To speed up the process, the 3+3 design is modified into a two-stage design. The first stage is to treat one patient at each dose level from the lowest until the first toxicity appears. Then starting at this dose, the traditional 3+3 design is used.

Although the 3+3 design is convenient to use, it has some problems that may affect the estimation of dose-finding (O'Quigley and Chevret, 1991), (O'Quigley and Shen, 1996), (Yin, 2012). First, the observation data is only associated with the current dose level, and other dose levels are not considered. Second, the 3+3 design has poor statistical properties. Third, the 3+3 design only works for a trial in which the target toxicity probability is smaller than 33%.

2.2.2 A+B Design

Lin and Shih (2001) extend the 3+3 design to A+B design that is more general in practice. The significant difference from the 3+3 design is that the cohort size

may not be three. The procedures are described as follows:

(1) Suppose that A patients are treated at the dose level i, and are observed for toxicity level.

(2) If less than C patients of the total A patients are observed for toxicity, we escalate to the higher dose level at $i + 1$.

(3) If the number of toxic patients is between C and D, we stay at the same dose level i where we will treat B more patients. If we observe more than E of A+B patients for toxicity, de-escalate to the next lower dose level $i - 1$, otherwise, we escalate to the next higher dose level $i + 1$.

(4) If the number of toxic patients are greater than D, de-escalate to the dose level $i - 1$.

As we know, the standard 3+3 design is a special case of A+B design when A and B are 3 and C, D and E are 1. For example, Lin and Shih (2001) considered the case of $A = 3, B = 6, C = 1, D = 2$ and $E = 2$.

2.2.3 Accelerated titration design

In a classic Phase I design, we start from the very low dose level to protect treated patients from toxicity, however, the determined MTD is always far behind the starting dose. So it takes a very long time and a large number of patients to involve the trial before entering the market for a new drug. Simon et al. (1997)

develop the accelerated titration design (ATD) to speed up the design. The first stage of the ATD is to treat one patient at each sequential dose level until observing toxicity. The second stage is the standard 3+3 design. We need two more patients to enter the trial, and after that three patients are treated starting from the previous dose level. This procedure provides the opportunity to save time at the beginning of the trial and to avoid a large number of patients treated at a very low dose level that is much lower than the effective dose.

2.2.4 Biased coin dose-finding method

The biased coin design (BCD) is another traditional method to determine the MTD. The BCD decides the number of patients in each treatment with tossing a biased coin. We start treating the first patient at the lowest dose or at the dose that we believe is safe and close to the target dosage. Select and fix a number r, say $r > 0.5$. If we observe toxicity for the currently treated patient, then the next patient will receive the lower dose with probability r, and the next higher dose with probability $1 - r$. Durham et al. (1997) apply a random walk design to extend the BCD. Suppose ϕ is the target toxicity probability. Assume $\phi \leq 0.5$. If the currently treated patient is toxic, de-escalate to the next lower dose level. If no toxicity is observed, escalate to the next higher dose level with probability $\frac{\phi}{1-\phi}$ or treat the current patient at the same dose with probability $\frac{1-2\phi}{1-\phi}$. Assume now $\phi > 0.5$. If we observe toxicity for the currently treated patient, then the next patient will receive the lower dose with probability $\frac{1-\phi}{\phi}$, or receive the same

dose level with probability $\frac{2\phi-1}{\phi}$. If we observe no toxicity, then the next patient will receive the higher dose. Stylianou and Flournoy (2002) develop the BCD to use the maximum likelihood method, weighted least-squares method and isotonic design.

2.2.5 Group up-and-down design

In the up-and-down design, the current cohort is treated at the dose level d_i. Suppose t_i patients are observed for toxicity. The cohort size is s. Assume two integers c_l and c_u, where $0 \leq c_l \leq c_u \leq s$. The next cohort of patients will receive the dose d_{i-1} if $t_j \geq c_u$, and the current dose is the lowest, then treat all patients with this dose. The next cohort of patients will receive the dose d_{i+1} if $t_j \leq c_l$, and the current dose is the highest, then treat all patients with this dose. The next cohort of patients will receive the dose d_i if $c_l \leq t_j \leq c_u$.

2.3 Parametric methods

2.3.1 Continual reassessment method

As we known, rule-based methods to dose-finding only follow some pre-specified rules, and we collect information based on the current dose, and have no information on other doses. To overcome the problems of nonparametric methods, we use the model-based methods called parametric methods in Phase I trials. First, we introduce the continual reassessment method (CRM), which is a commonly used

method in a dose-finding design.

O'Quigley et al. (1990) propose the continual reassessment method which connects the true toxicity probabilities $\pi_1, \pi_2, \cdots, \pi_J$ with prespecified toxicity probabilities $p_1, p_2, \cdots, p_J$ at each dose by a parametric model with an unknown parameter α. So the dose-finding decision making is based on the model that we define in the trial study. In general, we assume that the toxicity probability p_i is the function of dose level i and π_i is the function of p_i. Also p_i increases with the dose level. That is, $p_1 < p_2 < \cdots < p_J$, Let ϕ_T denote the target toxicity probability. The CRM assumes,

$$P(\text{toxicity at dose level } i) = \pi_i = p_i^{exp(\alpha)},$$

where α is the unknown parameter in the model (O'Quigley and Shen, 1996).

There are some modifications of CRM. For example, a logistic model or a hyperbolic tangent function can be used in the toxicity probability model. For example, we can use

$$\pi_i(\alpha, \beta) = \frac{\exp(\beta + \alpha d_i)}{1 + \exp(\beta + \alpha d_i)},$$

where α and β are unknown parameters, and d_i is the dose at level i after standardization.

Another model is the hyperbolic tangent function,

$$\pi_i(\alpha) = \left\{ \frac{\tanh(d_i) + 1}{2} \right\}^{\alpha} = \left\{ \frac{(e^{2d_i} - 1)/(e^{2d_i} + 1) + 1}{2} \right\}^{\alpha},$$

Then we are able to use frequentist or Bayesian approach to estimate the

MTD toxicity. In this chapter, we only focus on the Bayesian method. Suppose that n_i patients are treated at dose level i, for $i = 1, 2, \cdots, J$, and y_i patients over the total n_i patients experience toxicity. Assume that D is observed data set. The likelihood function is

$$L(D|\alpha) \propto \prod_{i=1}^{J} \{p_i^{\exp(\alpha)}\}^{y_i} \{1 - p_i^{\exp(\alpha)}\}^{n_i - y_i},$$

We assume a prior that follows a specified distribution denoted by $f(\alpha)$. By Bayes' theorem, the toxicity probability at dose level i can be investigated by the posterior mean denoted by $\hat{\pi}_i$ and given by

$$\hat{\pi}_i = \int p_i^{\exp(\alpha)} \frac{L(D|\alpha) f(\alpha)}{\int L(D|\alpha) f(\alpha) d\alpha} d\alpha.$$

After treating all the patients, we can obtain the posterior mean of the toxicity probability at each dose level, and find the dose level where the posterior mean is the closest to the target toxicity probability ϕ_T. Then this dose is our recommended dose level for the Phase II trial.

2.3.2 Bayesian model averaging CRM

In the CRM, we try to model the true toxicity probability. If our modeled toxicity probabilities are far from the true ones, the estimates may not be precise, and the design may not perform well. To find the MTD in the Phase I trial, Yin and Yuan (2009) apply multiple CRMs, each has a different prior, to model the true toxicity probability. Hoeting et al. (1999) propose a discrete prior probability

to each CRM and assign each CRM model a weight called a Bayeasian model averaging (BMA) procedure. In practice, we may assign more weight to the better fitted model. So the estimates of toxicity probabilities approach closest to the best fitted one all over CRMs.

The BMA-CRM design uses multiple CRM models. Suppose M_k denotes the k^{th} CRM model related to toxicity probability set $(p_{k1}, p_{k2}, \cdots, p_{kJ})$, where $k = 1, 2, \cdots, K$, and K is the total number of CRM models.Then the toxicity probability at d_i is

$$\pi_{ki}(\alpha_k) = p_{ki}^{\exp(\alpha_k)},$$

where $i = 1, 2, \cdots, J$, J is the total number of dose levels, and α_k is the unknown parameter related to CRM model M_k (Yin, 2012). Suppose that y_i patients who are treated over the total n_i patients experience toxicity. Assume that D is the observed data set. The likelihood function of CRM model M_k is

$$L(D|\alpha_k, M_k) \propto \prod_{i=1}^{J} \{p_{ki}^{\exp(\alpha)}\}^{y_i} \{1 - p_{ki}^{\exp(\alpha)}\}^{n_i - y_i}.$$

Then assume that a prior follows a specified distribution, denoted by $f(\alpha_k|M_k)$, in the model M_k. This implies the likelihood function of the model M_k is

$$L(D|M_k) = \int L(D|\alpha_k, M_k) f(\alpha_k|M_k) d\alpha_k,$$

and the posterior model probability for M_k is

$$P(M_k|D) = \frac{L(D|M_k)P(M_k)}{\sum_{i=1}^{K} L(D|M_i)P(M_i)}.$$

Finally, the toxicity probability at dose level j is estimated by the Bayesian model averaging method as

$$\bar{\pi}_j = \sum_{k=1}^{K} \hat{\pi}_{kj} P(M_k|D), j = 1, 2, \cdots, J.$$

Here,

$$\hat{\pi}_{kj} = \int p_{kj}^{\exp(\alpha_k)} \frac{L(D|\alpha_k, M_k) f(\alpha_k|M_k)}{\int L(D|\alpha_k, M_k) f(\alpha_k|M_k) d\alpha_k} d\alpha_k$$

is the posterior mean of the toxicity probability at dose level j, under the assumption of model M_k.

So the Bayesian model averaging estimate of the toxicity probability is a weighted average of the posterior means, where the weight is given by $P(M_k|D)$. After treating patients at dose level j, the decision of whether to escalate or de-escalate the dose depends on the value of $\bar{\pi}_j$.

An important issue is to develop an algorithm to find MTD that is closest to the prescribed target toxicity probability.

2.3.3 Escalation with overdose control

Babb et al. (1998) develop the escalation with overdose control (EWOC) design to protect patients from overdose treatment. Assume $y_i = 1$ if the patient is toxic, and $y_i = 0$ otherwise. Then the toxicity probability of dose level i is the function F of dose level. Define

$$P(\text{toxicity probability of dose } i) = P(y_i = 1|dose = x_i) = F(\alpha + \beta x_i),$$

where α and β are unknown parameters. Suppose there are n_i patients involved in the trial study, and the observed data set is $Y = \{y_1, \cdots, y_n\}$. The likelihood function is given by

$$L(Y|\alpha, \beta) = \prod_{i=1}^{n} \{F(\alpha + \beta x_i)\}^{y_i} \{1 - F(\alpha + \beta x_i)\}^{1-y_i}.$$

Suppose M and ϕ_T denote the MTD and the target toxicity probability respectively. Then x_0 is assumed to be the lowest dose level. Then

$$\phi_T = P(y_i = 1|dose = M) = F(\alpha + \beta M),$$

and

$$\pi_0 = P(y_i = 1|dose = x_0) = F(\alpha + \beta x_0).$$

So we can calculate α and β as follows:

$$\beta = \frac{F^{-1}(\phi_T) - F^{-1}(\pi_0)}{M - x_0},$$
$$\alpha = F^{-1}(\pi_0) - \beta x_0.$$

We can assume that $F^{-1}(x)$ is exponential, logistic or hyperbolic tangent function. After assuming the prior distributions of M and π_0, we can get the joint posterior distribution of M and π_0. Integrating out π_0, the marginal distribution of M, denoted as $G(x|Y)$, can be used to find the next dose level. We define,

$$G(x|Y) = P(M \leq x|Y).$$

In the EWOC design, the minimal posterior expected loss is used to determine the appropriate dose,

$$\int L_\gamma(x, M) dG(M|Y),$$

where γ is a pre-specified threshold, and $L_\gamma(x, M)$ is an asymmetric loss function (Yin, 2012).

2.3.4 Bayesian hybrid design

The Bayesian hybrid design can switch from the non-parametric methods to model-based parametric methods. We can take advantages from both of them to determine the MTD (Yuan and Yin, 2011). If the information is enough to determine whether the current dose level is below or above the MTD, we can use the non-parametric method. If the information from current dose level is not enough, we introduce the model-based parametric method to determine escalation or de-escalation of the dose level.

Suppose π_i is the toxicity probability at the current dose level i, and we use three hypotheses to determine where dose π_i locates relative to the target toxicity probability ϕ_T. Define the following hypotheses:

$H_1 : \pi_i \leq \phi_T - \delta,$

$H_2 : \phi_T - \delta \leq \pi_i \leq \phi_T + \delta,$

$H_3 : \pi_i > \phi_T + \delta,$

where δ is a toxicity tolerance margin (say, 0.33).

Assume a conditional prior distribution of π_i given each hypothesis follows a uniform distribution, that is,

$$\pi_i|H_1 \sim Unif(0, \phi_T - \delta),$$
$$\pi_i|H_2 \sim Unif(\phi_T - \delta, \phi_T + \delta),$$
$$\pi_i|H_3 \sim Unif(\phi_T + \delta, 1).$$

Let y_i denote the number of patients who experience toxicity. The posterior distribution of each hypothesis is

$$\begin{aligned} P(H_k|y_i) &= \frac{P(H_k)}{P(H_1)BF_{1,k} + P(H_2)BF_{2,k} + P(H_3)BF_{3,k}}, \\ BF_{m,k} &= P(y_i|H_m)/P(y_i|H_k), \text{for } m = 1, 2, 3. \end{aligned}$$

The marginal distribution of y_i given H_1 is

$$P(y_i|H_1) = \int_0^{\phi_T - \delta} \binom{n_i}{y_i} \pi_i^{y_i}(1-\pi_i)^{n_i - y_i} \frac{1}{\phi_T - \delta} d\pi_i.$$

Similarly, we can obtain the marginal distribution of y_i given H_2 and H_3. Jeffreys (1961) proposes that if $P(H_1|y_i) > 0.61$, the dose level i is below the MTD, and we need to escalate to the next higher doze level $i + 1$. If $P(H_2|y_i) > 0.61$, we will stay at the same dose level. If $P(H_3|y_i) > 0.61$, we need to de-escalate to the next lower dose level $i - 1$.

2.4 Summary

In this chapter, we have reviewed basic ideas of commonly used nonparametric and parametric designs of Phase I clinical trials. To summarize, the goal of Phase I clinical trial is to assess the toxicity of the new drug and identify the maximum tolerable dose. Details of the continual reassessment method are given and will be used to extend the method in the next two chapters.

Chapter 3

A new design with immediate toxicity

3.1 Introduction

row In this chapter, we introduce a new parametric design of dose finding in Phase I clinical trials by the continual reassessment method (CRM). The true toxicity probability of the drug is connected with the pre-specified toxicity probability through the CRM at each dose level. In this clinical trial, patients are assigned to the dose level most possibly closest to the target, depending on previous toxicity probabilities. The process continues until the MTD is determined or the maximum sample size is reached.

In general, let $d_i, i = 1, 2, \cdots, K$, be the dose levels and p_i be the pre-specified toxicity probability at dose level d_i. We assume, as commonly used in the literature, the toxicity probability is monotonically increasing with the dose level. This means $p_1 < p_2 < \cdots < p_K$ at dose level $d_i, i = 1, 2, \cdots, K$. Let ϕ_T be the target toxicity

probability. In a dose finding Phase I clinical trial, a cohort of patients (usually of size three or of size one) is sequentially assigned to a chosen dose level, and all patients in the same cohort receive the same dose level. Finally we observe a binary toxicity outcome Y_j for each patient j, where $Y_j = 1$ if the toxicity is observed and $Y_j = 0$ if no toxicity is observed. The CRM assumes a dose-toxicity model by defining

$$Pr(\text{toxicity at dose level } i) = \pi_i(\alpha) = {p_i}^{exp(\alpha)},$$

where α is an unknown parameter (O'Quigley and Shen, 1996).

The CRM initially introduced in O'Quigley et al. (1990) has been extended and improved in many directions, often in terms of different dose-toxicity models. For example, O'Quigley and Shen (1996) introduces a one-parameter logistic function

$$\pi_i(\alpha) = \frac{exp(\beta + \alpha d_i)}{1 + exp(\beta + \alpha d_i)},$$

where β is a fixed constant and d_i is the standardized dose level i. In particular, Yin (2012) applies a fixed constant β of -3. A hyperbolic tangent function,

$$\pi_i(\alpha) = \left\{\frac{\tanh(d_i) + 1}{2}\right\}^{\alpha} = \left\{\frac{(e^{2d_i} - 1)/(e^{2d_i} + 1) + 1}{2}\right\}^{\alpha},$$

was also introduced in O'Quigley et al. (1990).

In this thesis, we introduce a new parametric model for the dose-toxicity probability function, given by

$$\pi_i(\alpha) = \frac{2\Phi(\beta + \alpha d_i)}{1 + \Phi(\beta + \alpha d_i)},$$

where $\Phi(x) = \Phi(x, \mu, \sigma^2)$ is the cumulative distribution function of the normal distribution with mean μ and variance σ^2. The rationale is that this function of d_i is increasing and changes concavity and their values are between zero and one. Furthermore, $\Phi(x)$ can also be the cumulative distribution function of other distributions, which we will investigate in the future. To compare our new design with the current designs, we set $\beta = -3$ in our simulations.

To guarantee an increasing relationship between dose level and toxicity probability, the parameter α in the new model is restricted to be positive. This can be checked by the first derivative of the true toxicity probability function.

The CRM assumes that the parameter α is random and follows a prior distribution $f(\alpha)$ and uses a Bayesian approach to obtain the posterior mean to estimate the toxicity probability at dose level i. Cheung (2011) proposes the normal prior distribution, that is, α follows a normal distribution with mean μ_α, and variance σ_α^2. Suppose that we have a prior belief that the dose level i^* is the MTD, and set $p_{i^*} = \theta$, where θ is the target probability, say $\theta = 0.25$. We assume that $p_1 < p_2 < \cdots < p_{i^*} < \cdots < p_K$, which means the toxicity probability is a strictly increasing function in dose level. For the logistic function, the backward substitution gives us

$$d_i = \frac{logit(p_{i^*}) - \beta}{exp(\mu_\alpha)},$$

where $logit(p) = log(p/(1-p))$. Then we have

$$F(d_i, \alpha) = \frac{exp[\beta + exp(\alpha - \mu_\alpha)(logit(p_{i^*}) - \beta)]}{1 + exp[\beta + exp(\alpha - \mu_\alpha)(logit(p_{i^*}) - \beta)]},$$

where $F(d_i, \alpha)$ is the toxicity probability, only depending on $\alpha - u_\alpha$, which is normally distributed with mean zero. So we may set $\mu_\alpha = 0$ to simplify the computation. row The CRM starts to treat the first cohort of patients at the prior MTD i^*. Note that the starting dose has a toxicity probability θ. Each increasing dose is determined by previous observations and is obtained by likelihood function and posterior mean of the toxicity probability. At the model-based MTD estimation, we can find the dose level whose posterior mean is closest to target toxicity probability.

Let D be the observed information, then the likelihood function is given by

$$L(D|\alpha, \beta) \propto \prod_{i=1}^{K} \{\pi_i\}^{y_i} \{1 - \pi_i\}^{1-y_i}.$$

Using our definition of π_i, the likelihood function becomes

$$L(D|\alpha) \propto \prod_{i=1}^{K} \left\{ \frac{2\Phi(\beta + \alpha d_i)}{1 + \Phi(\beta + \alpha d_i)} \right\}^{y_i} \left\{ 1 - \frac{2\Phi(\beta + \alpha d_i)}{1 + \Phi(\beta + \alpha d_i)} \right\}^{1-y_i}.$$

We assume α is positive. To compare with existing methods, we assume $\beta = -3$. By the Bayes' theorem, the toxicity probability at dose level d_i is estimated to be

$$\hat{\pi}_i = \int \frac{2\Phi(\beta + \alpha d_i)}{1 + \Phi(\beta + \alpha d_i)} \frac{L(D|\alpha) f(\alpha)}{\int L(D|\alpha) f(\alpha) d\alpha} d\alpha,$$

where $\hat{\pi}_i$ is the posterior mean of the toxicity probability at dose d_i. After each cohort of patients is treated, we collect all the toxicity data and calculate the posterior means of the toxicity probabilities at all the dose levels, say $\hat{\pi}_1, \hat{\pi}_2, \cdots, \hat{\pi}_K$. The dose whose toxicity probability is closest to the target Φ_T is recommended to

the next cohort of patients. The trial terminates when the toxicity probability converges, then we determine this dose level as the MTD.

The dose finding procedure by CRM follows the following rules (Yin, 2012):

(1) We treat the first cohort of patients at the starting dose.

(2) Let the current dose level be i^{cur}, and denote the target toxicity probability as Φ_T. We can calculate the posterior means of the toxicity probabilities for all the observations, that is, $\hat{\pi}_1, \hat{\pi}_2, \cdots, \hat{\pi}_K$. We obtain the dose level i^* whose toxicity probability closest to Φ_T,

$$i^* = argmin|\hat{\pi}_i - \Phi_T|,$$

If $i^{cur} > i*$, de-escalate to the next lower level, and if $i^{cur} < i*$, escalate to the next higher level, otherwise, keep it at the same dose level.

(3) We determine the dose with the toxicity probability closest to Φ_T as the MTD when the maximum sample size is collected.

3.2 A new design

In this chapter, we extend the above models by introducing a new dose-toxicity model. This was motivated by my graduate course in nonlinear models where the the cumulative distribution function of the normal distribution is used.

Specifically, we introduce the model

$$\pi_i = \pi_i(d_i, \alpha, \beta) = \frac{2\Phi(\beta + \alpha d_i)}{1 + \Phi(\beta + \alpha d_i)} = 2 - 2(1 + \Phi(\beta + \alpha d_i))^{-1},$$

where Φ is the cumulative distribution function (CDF) of the normal distribution $N(\mu, \sigma^2)$, not necessarily the standard normal distribution. The reason for choosing a general normal distribution is that we can model different shapes of the increasing dose-toxicity relationship.

We can check the increasing monotonicity of the dose-toxicity function by its first derivation. Using calculus, for given μ and σ, we have

$$\frac{d\pi_i}{dd_i} = \frac{2\alpha\phi(\beta + \alpha d_i)}{(1 + \Phi(\beta + \alpha d_i))^2},$$

where ϕ is the probability density function (PDF) of the normal distribution and ranges from zero to positive infinite. To ensure a positive first derivative, we only need that α is positive.

From this first derivative, we see that the dose-toxicity function is increasing in dosage as long as the parameter α takes positive values.

We understand the concavity of the dose-toxicity function by checking its second derivation, which is give by

$$\begin{aligned}\frac{d^2\pi_i}{dd_i^2} &= \frac{2\alpha^2\phi(\beta + \alpha d_i)'(1 + \Phi(\beta + \alpha d_i))^2 - 4\alpha^2\phi(\beta + \alpha d_i)^2(1 + \Phi(\beta + \alpha d_i))}{(1 + \Phi(\beta + \alpha d_i))^4} \\ &= \frac{2\alpha^2\phi(\beta + \alpha d_i)'(1 + \Phi(\beta + \alpha d_i)) - 4\alpha^2\phi(\beta + \alpha d_i)^2}{(1 + \Phi(\beta + \alpha d_i))^3}.\end{aligned}$$

From this second derivative, we see that the dose-toxicity curve π_i concaves up in d_i if $\frac{d^2\pi_i}{dd_i^2} > 0$ and concaves down in d_i if $\frac{d^2\pi_i}{dd_i^2} < 0$. However there is no closed

form for the CDF Φ, so we see intuitively that for some values of d_i, we have $\frac{d^2\pi_i}{dd_i^2} > 0$, and for some other values of d_i, we have $\frac{d^2\pi_i}{dd_i^2} < 0$. The second derivative $\frac{d^2\pi_i}{dd_i^2}$ is positive if

$$(1 + \Phi)\phi' - 2\phi^2 \geq 0.$$

This is true if $\frac{\phi'}{\phi} \geq \frac{2\phi}{1+\Phi}$. That is, if $(ln\phi)' \geq (2ln(1 + \Phi))'$, the second derivative $\frac{d^2\pi_i}{dd_i^2}$ is positive. Because a logarithmic function is a one-to-one and monotonically increasing function, this is true if $\frac{\phi}{(1+\Phi)^2}$ is increasing in d_i. To check the concavity of pi_i in d_i, we need just to check this condition which is true with the help of R.

Using the statistical software R, we draw the graphs of the dose-toxicity relation for different values of μ and σ^2. To maintain readability, we have included in Appendix B (Figures B.1 to B.16) the graphs of the new function and the three functions used in the literature, for different values of the parameters. From these graphs, we see that our new model can depict a variety of dose-toxicity relations, just like the functions discussed in Section 3.1 (that is, the power function, the logistic model and the hyperbolic tangent model). However we hope our model performs better because we can choose different values of μ and σ^2 to adjust the shape of the dose-toxicity curve.

3.3 Simulation for the new design

For the purpose of illustration, we set $\beta = -3$. In this section, we carry out some simulation studies to assess the performance of our new design, using the following criteria: (1) the convergence rate of the dose-selection process, (2) average proportion of toxic patients. The prior distribution of the unknown parameter is taken to be the beta distribution $beta(2, 2)$. The first criterion tells us that the faster the convergence is, the less patients we have in the trial. This makes the trial more ethical because we subject less patients to toxicity overall. The second criterion tells us how many patients are toxic in a particular trial, so with the same sample size, the trial with less toxic patients is more ethical.

We assume the unknown parameter α follows a beta prior distribution with the probability function, and its probability density function is given by $f(x) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1}(1-x)^{s-1}$, where $0 < x < 1$. This ensures that α is positive and the dose-toxicity probability function is increasing. Although the posterior distribution has no closed form, its support is given by $(0, 1)$ and α is always positive. We use R programming to derive the posterior distribution and sample from this distribution. The probability of toxicity is estimated after treating each patient, and the trial is stopped if the difference of two consecutive estimated toxicity probabilities of the chosen doses is less than 0.005.

For example, corresponding to the dose trace illustrated on page B-19, the associated table of prior and posterior mean toxicity probabilities of our new

model is given on page B-18. The first row of this table gives the prior mean toxicity probabilities at all doses. The trial started with the lowest dose 1, as given by the algorithm. From page B-19, we see that patient 1 is not toxic. So we derive the likelihood, update the posterior distribution and calculate the posterior mean toxicity probabilities at all doses. Then for the second patient, we apply the dose whose posterior mean toxicity probability is below 0.33 but closest to 0.33. From this table, we see that dose 4 is selected. From the results on page B-19, we see that patient 2 is not toxic. We derive the likelihood again, update the posterior distribution and calculate the posterior mean toxicity probabilities at all doses. The results are given in the third row of the table. For the third patient, we select the dose with the posterior mean toxicity probability that is below 0.33 but closest to 0.33. Dose 4 is selected again, the patient is not toxic from page B-19. We derive the likelihood again, update the posterior distribution and calculate the posterior mean toxicity probabilities at all doses. The results are given in the fourth row of the table. For the fourth patient, we select the dose with the posterior mean toxicity probability that is below 0.33 but closest to 0.33. This time, after observing two non-toxic results at dose 4, we escalate to dose 5. This seems intuitive and reasonable. Unfortunately after treatment, the fourth patient is toxic. So we would expect to de-escalate to dose 4 again. This is confirmed by calculation. After observing a toxic response, we derive the likelihood, update the posterior distribution and calculate the posterior mean toxicity probabilities at all doses. The results are given in the fifth row of the

table. From this row, we see that dose 4 has a posterior mean toxicity probability that is below 0.33 but still closest to 0.33. The fifth patient is treated at dose 4. But again the patient is toxic. We expect the next patient be treated at the next lower dose 3. This is confirmed by calculation again. After observing a toxic response, we derive the likelihood, update the posterior distribution and calculate the posterior mean toxicity probabilities at all doses. The results are given in the sixth row of the table. From this row, we see that dose 3 has a posterior mean toxicity probability that is below 0.33 but still closest to 0.33. This and the next 4 patients (i.e., a total of 5 patients) are treated at dose 3 and all are non-toxic. Intuitively for the 11^{th} patient, we would escalate the dose 4. This is confirmed by calculation. We derive the likelihood, update the posterior distribution and calculate the posterior mean toxicity probabilities at all doses for the 11^{th} patient. The results are given in the eleventh row of the table. From this row, we see that dose 4 has a posterior mean toxicity probability that is below 0.33 but still closest to 0.33. After treatment, a toxic response is observed. As a result, we de-escalate to dose 3 for the next patient. This process repeats. The next 5 patients are treated at dose 3 and all are non-toxic. So we escalate to dose 4 again, but a toxic response is observed. We de-escalate to dose 3 again for the next 5 patients, and all are observed with non-toxic responses. We then escalate to dose 4, but again a toxic response is observed. After this, all future patients are treated at dose 3, and the 26^{th} patient was toxic. Nevertheless, dose 3 is identified as the MTD after convergence. The dose selection process stopped after treating 30 patients.

Of the 30 patients treated, a total of 6 patients are toxic, or 20%. The sample mean and variance of the toxicity probabilities of the 30 treated patients are respectively 0.204 and 0.004. Finally, at the time of stopping, the final estimated dose probabilities are 0.0148 (at dose 1), 0.0604 (at dose 2), 0.1763 (at dose 3), 0.3680 (at dose 4), 0.5821 (at dose 5), and 0.7591 (at dose 6). Dose 3 is correctly identified as the MTD, the simulated proportion of toxic patients is slightly higher than the estimated toxicity probability at dose 3, and the sample mean toxicity probability of the 30 patients in the trial is slightly higher than the identified dose probability at MTD.

For another dose trace illustrated on page B-20, the dose selection process stopped after treating 38 patients. Of the 38 patients treated, a total of 10 patients are toxic, or 26.3%. The sample mean and variance of the toxicity probabilities of the 38 treated patients are respectively 0.232 and 0.002. At the time of stopping, the final estimated dose probabilities are 0.0168 (at dose 1), 0.0736 (at dose 2), 0.2180 (at dose 3), 0.4431 (at dose 4), 0.6702 (at dose 5), and 0.8353 (at dose 6). Dose 3 is correctly identified as the MTD, the simulated proportion of toxic patients is slightly higher than the estimated toxicity probability at dose 3, and the sample mean toxicity probability of all 38 treated patients is also slightly higher than the identified dose probability at MTD. Except for the first 7 patients, all other patients received dose level 3 for the treatment.

We also repeat the simulation for the case of 10 dose levels. To adjust the

dose toxicity probabilities, we have changed the toxicity probability function to

$$\pi_i(\alpha) = \frac{2\Phi(-5 + \alpha d_i)}{1 + \Phi(-5 + \alpha d_i)}.$$

Here we have changed -3 to -5 because this way the transformed dose $-5 + \alpha d_i$, where $d_i = i, i = 1, 2, \cdots, 10$, ranges from -5 to 5, since $0 \leq \alpha \leq 1$. Based on this new toxicity probability function, the toxicity probabilities at all 10 dose levels seem reasonable. For example, corresponding to the dose trace with 10 doses illustrated on page B-21, the dose selection process stopped after treating 60 patients. Of the 60 patients treated, a total of 10 patients are toxic, or 16.7%. The mean and variance of the toxicity probabilities of all 60 treated patients are respectively 0.214 and 0.002. Finally, at the time of stopping, the final estimated dose probabilities from dose 1 to 10 are 0.00001, 0.0002, 0.0021, 0.0151, 0.0712, 0.2206, 0.4589, 0.6973, 0.8631, 0.9502 respectively. Dose 6 is correctly determined as the MTD, the simulated proportion of toxic patients is lower than the estimated toxicity probability at dose 6, and the sample mean toxicity probability of all 60 treated patients is also lower than the estimated toxicity probability at MTD. It is interesting to observe that after starting the first patient with the lowest dose and no toxicity, the second patient is treated at dose 7. The second patient was toxic so we de-escalated to dose 5, for the next 20 patients. At the beginning , 2 toxicities were observed at dose 5 (for the 5^{th} and 8^{th} patients). However for all other patients at dose 5, non-toxicity was observed and so we escalate to dose 6 for the next 38 patients. Among these 38 patients, 7 toxicities were observed and the algorithm converged at dose 6.

In the dose trace with 10 dose levels illustrated on page B-22, the dose selection process stopped after treating 44 patients. Of these 44 patients treated, a total of 12 patients are toxic, or 27.3%. The sample mean and variance of the toxicity probabilities of all 44 treated patients are respectively 0.232 and 0.006. Finally, at the time of stopping, the final estimated dose probabilities from dose 1 to 10 are 0.0000001, 0.000006, 0.0002, 0.0024, 0.0215, 0.1119, 0.3350, 0.6230, 0.8400, 0.9494 respectively. Dose 6 is correctly determined as the MTD, the simulated proportion of toxic patients is higher than the estimated toxicity probability at dose 6, and the sample mean toxicity probability of all 44 treated patients is also higher than the estimated toxicity probability at MTD. The dose trace is interesting. We start with the lowest dose and the observation is non-toxic. The second treatment is escalated to dose 9 but the observation was toxic. We then de-escalate to dose 6. After observing a non-toxic response, we escalate again, to dose 7 for the next 2 patients. However 1 toxic response was observed and so we de-escalated to dose 6 again, for the next 3 patients, who were all non-toxic. So we escalate to dose 7 again, but the observation was toxic and so we de-escalate again to dose 6, for the next 6 patients. All patients were non-toxic and so we escalate again to dose 7, for the next 24 patients. After observing 8 toxicities, we de-escalated to dose 6 and the treatment was non-toxic. We treated 2 more patients at dose 7, and the last one was toxic. We treated the last 2 of the 44 patients at dose 6, observed no toxicity, and the algorithm converged.

For the above simulations, we used the $Beta(2, 2)$ distribution. This distri-

bution is symmetric about 0.5 and gives equal weights to both large and small values in $(0, 1)$. We have extended this to a general beta distribution $Beta(a, b)$. When the first parameter a is smaller than b, we give more weights to small values in $(0, 1)$ than large values. This means that both the prior mean and posterior mean of the toxicity probability function $\pi_i(\alpha)$ tend to be small for all doses $i = 1, 2, \cdots, 6$. Because of this, the MTD is likely to be at dose 6. This has been observed from page B-23 and page B-24, where the proportion of dose 6 as MTD has increased. At the same time, because all toxicity probabilities are small and dose 6 is the MTD in most times, we expect relatively fast convergence. This is also observed on page B-25 and page B-26.

On the other hand, we simulated with the $Beta(a, b)$ distribution when a is larger than b. In this case, we give more weights to large values in $(0, 1)$, and therefore, most or all the prior and posterior mean toxicity probabilities are large, and in almost all cases larger than 0.33. As a result, no MTD were selected according to our criteria.

We also simulated the trial using the uniform distribution over $(0, 1)$ as the prior distribution. Because we do not favour either small or large values in $(0, 1)$, we have observed that both dose 1 and dose 6 are very likely to be the MTD, and the convergence time should be quick too. These have been observed on page B-27 and page B-28.

3.4 Simulation comparison with other designs

In this section, we carry out simulation studies to compare the performance of our new design with existing designs. The comparison is based on the following three criteria: (1) the rate of convergence of the design, (2) the average number of toxic patients in the trial, and (3) the average toxicity of all patients treated in the trial. All simulations are based on 1000 replications. All simulations use the $Beta(a, b)$ prior where $a = b = 2$ which are very good as we have observed in Section 3.3. The stopping rule is the same for all designs. That is, we stop each design when the absolute difference of the toxicity probabilities of the last two chosen doses is less than 0.005. The dose selection rule is also the same. That is, for each patient, the dose selected is the one whose toxicity probability is less than 0.33 but at the same time closest to 0.33.

The motivation of using these three criteria is as follows. The rate of convergence tells us how many patients we treat, on average, before we stop the trial. The intuition is that the less patients we have in the trial, the less number of patients are subject to toxicity. For a fixed number of patients in the trial, the less number of toxic patients, the better the design. So we wish to estimate the average number of toxic patients for each design. Finally, we estimate the average toxicity probability of all patients in the trial. We hope that the lower this average, the better the design, because this average tells us the overall level of toxicity of all patients treated in the trial.

For our new design, the dose toxicity probability function is given by $\pi_i(\alpha) = \frac{2\Phi(-3+\alpha d_i,\mu,\sigma^2)}{1+\Phi(-3+\alpha d_i,\mu,\sigma^2)}$, where α is the unknown parameter following the $Beta(2,2)$ prior distribution, μ is the mean of the normal distribution and σ^2 is the variance of the normal distribution, and $d_i = i, i = 1, 2, \cdots, 6$, is the dose. The dose toxicity probability function of the logistic design is given by $\pi_i(\alpha) = \frac{exp(-3+\alpha d_i)}{1+exp(-3+\alpha d_i)}$, where α is the unknown parameter following the $Beta(2,2)$ prior distribution and $d_i = i, i = 1, 2, \cdots, 6$, is the dose. The dose toxicity probability function of the hyperbolic tangent design design is given by $\pi_i(\alpha) = \left[\frac{tanh(-3+d_i)+2}{2}\right]^{\alpha}$, where α is the unknown parameter following the $Beta(2,2)$ prior distribution and $d_i = i, i = 1, 2, \cdots, 6$, is the dose.

Simulation results are summarized in Appendix C. For our new design, we simulated the cases of $N(0, 0.5)$, $N(0, 1)$, $N(0, 1.5)$, $N(0, 2)$ and $N(0, 3)$. From Table C.1 on page C-2, we see our new design performs the best. The average stopping times for our new design (19.33 for $\sigma^2 = 0.5$, 24.13 for $\sigma^2 = 1$, 32.09 for $\sigma^2 = 1.5$, etc.) are lower than those for the logistic (33.96) and hyperbolic tangent (42.08) designs, so our new design is efficient. This means that on average, small sample sizes are needed to complete Phase I clinical trials using our new design. This is both logistically required and ethically desirable. Our new design with $\sigma^2 = 1$ is easy to apply because it is the standard normal distribution. For this design, the average proportion of toxic patients in the trial is 24.8%, which is lower than the average proportion of 27.7% for the logistic design and 26.6% for the hyperbolic tangent design. This implies that on average, we subject less

patients in the trial to toxicity if we use our new design. This means that our new design is more ethical than both logistic and hyperbolic tangent designs. Finally the average toxicity probability for all patients is 20.6% for our new design with $N(0, 1)$, that is lower than 25.5% for the logistic design and 28.6% for the hyperbolic tangent design. This also shows that our new design is more ethical.

It is worth pointing out that for our new model with $\sigma^2 = 3$, all prior mean toxicity probabilities are more than 0.33 and therefore the trials were inconclusive and so the variances observed are 0 in Table C.1.

Table C.2 on page E-2 summarizes the simulated average toxicity probabilities at all dose levels under different models. These estimated probabilities show that the new design with the standard normal distribution seems to be a very reasonable design for the dose-toxicity relationship. The estimated average toxicity probabilities at different doses are 0.014 for dose 1, 0.063 for dose 2, 0.176 for dose 3, 0.327 for dose 4, 0.470 for dose 5 and 0.583 for dose 6. In this case, does 4 is recommended as MTD for a Phase II clinical trial.

3.5 Summary

This chapter is exclusively focused on simulation studies of the new design alone and also in comparison with existing designs. One advantage of our new design is we can adjust the variance of the normal distribution in order to model different possible patterns of the dose-toxicity relationship. This can normally not be

achieved by the three functions currently used in the literature. Simulation studies show that the new design performs much better than the three standard designs with respect to different criteria including the convergency rate and average proportion of toxic patients. Finally the MTD is identified together with its probability.

Chapter 4

A new design with late-onset toxicity

4.1 Introduction

In Chapter 3, we introduced a new design of dose finding in Phase I clinical trials when we assumed immediately observed toxicity. However it may happen that the toxicity cannot be observed immediately after the treatment. Instead it may be late-onset.

In this chapter, we carry out simulation studies when we assume the toxicity after treatment is delayed and so late-onset. Specifically we assume the time to toxicity follows a geometric distribution. Therefore when a particular patient is treated, the toxicity responses from previously treated patients may not be available. Instead all we know is some of these patients are not toxic at the time of treatment. Therefore we require the use of survival analysis to modify the likelihood function. The idea of using survival analysis came from taking the

Lifetime Data Analysis course from Dr. Po Yang.

4.2 A new design with delayed toxicities

We still use the following same model describing the dose-toxicity relationship:

$$\pi_i = \pi_i(d_i, \alpha, \beta) = \frac{2\Phi(\beta + \alpha d_i)}{1 + \Phi(\beta + \alpha d_i)} = 2 - 2(1 + \Phi(\beta + \alpha d_i))^{-1},$$

where Φ is the cumulative distribution function (CDF) of the normal distribution $N(\mu, \sigma^2)$, not necessarily the standard normal distribution.

We assume the time T to toxicity follows the geometric distribution with rate p. The probability mass function of T is given by $p(t) = p(1-p)^t, t = 0, 1, 2, \cdots$. In our simulations, the rate p will take values $0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$. Because the expected value of T is $\frac{1-p}{p}$, so the expected time to toxicity is respectively $9, 4, 2.33, 1.5, 1, 0.67, 0.43, 0.25, 0.11$ when p is $0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$. Our simulations consider different possible cases.

Due to late-onset toxicities, the likelihood function needs modification. Type I censoring Maximum Likelihood function at time m is

$$\prod_{l=1}^{m-1} [\pi_{i_l}(1 - \pi_{i_l})^{t_{i_l,l}}]^{\delta_{i_l,l}} [(1 - \pi_{i_l})^{m-l}]^{1-\delta_{i_l,l}},$$

where i_l is the dose for patient treated at time $l = 1, 2, \cdots, m-1$, $t_{i_l,l}$ is the time to toxicity for patient treated at time $l = 1, 2, \cdots, m-1$ with dose i_l, and

$$\delta_{i_l,l} = I_{\{t_{i_l,l} \le m-l\}} = \begin{cases} 1 & \text{if toxicity is observed,} \\ 0 & \text{if toxicity is censored,} \end{cases}$$

is an indicator. We also assume α is positive. To compare with existing methods, we set $\beta = -3$. By the Bayes' theorem, the toxicity probability at dose d_i is estimated to be

$$\hat{\pi}_i = \int \frac{2\Phi(\beta + \alpha d_i)}{1 + \Phi(\beta + \alpha d_i)} \frac{L(D|\alpha) f(\alpha)}{\int L(D|\alpha) f(\alpha) d\alpha} d\alpha,$$

where $\hat{\pi}_i$ is the posterior mean of the toxicity probability at dose d_i.

After each cohort of patients is treated, we collect all the complete and censored toxicity data and calculate the posterior means of the toxicity probabilities at all the dose levels, say $\hat{\pi}_1, \hat{\pi}_2, \cdots, \hat{\pi}_K$. The dose whose toxicity probability is closest to the target Φ_T is the dose that is recommended to the next cohort of patients. The trial terminates when the toxicity probability converges, and we determine this dose level as the MTD.

4.3 Simulation for the new design with late-onset toxicity

For the purpose of illustration, we set $\beta = -3$. In this section, we carry out some simulation studies to assess the performance of our new design when the toxicity is late-onset. That is, when the next patient is to be treated, the toxicity of a previously treated patient is delayed and may not be observable. We use the following criteria: (1) the convergence rate of the dose-selection process, and (2) average proportion of toxic patients. The prior distribution of the unknown parameter is taken to be the beta distribution $beta(2, 2)$. The first criterion tells

us that the faster the convergence is, the less patients we have in the trial. This makes the trial more ethical because we subject less patients to toxicity overall. The second criterion tells us how many patients are toxic in a particular trial, so with the same sample size, the trial with less toxic patients is more ethical.

We assume the unknown parameter α follows a beta prior distribution with the probability function, and its probability density function is given by $f(x) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1}(1-x)^{s-1}$, where $0 < x < 1$. This ensures that α is positive and the dose-toxicity probability function is increasing. Although the posterior distribution has no closed form, its support is given by $(0, 1)$ and α is always positive. We use R programming to derive the posterior distribution and sample from this distribution. The probability of toxicity is estimated after treating each patient, and the trial is stopped if the difference of two consecutive estimated toxicity probabilities of the chosen doses is less than 0.005.

The late-onset toxicity is simulated as follows, in two stages. After treatment, we observed that the patient is either non-toxic or toxic. If the patient is non-toxic, then we use this endpoint to update our likelihood function. If the patient is toxic, say at a dose with toxicity probability p, then we assume that the time to toxicity follows the geometric distribution $P(X = x) = p(1-p)^x, x = 0, 1, 2, \cdots$. This is a very reasonable discrete distribution to describe time to the event of toxicity. The random time to toxicity is generated by R using the rate $1/p$. Suppose that the generated time to toxicity is x and the patient is treated at time $i < n$ and

the current patient is treated at time n. If $x \leq n - i$, toxicity is observed and the contribution to the likelihood is $p(1-p)^x$. If $x > n - i$, toxicity is censored and the contribution to the likelihood is $(1-p)^{n-i+1}$.

For example, corresponding to the dose trace illustrated on page D-3, the associated table of prior and posterior mean toxicity probabilities of our new model is given on page D-2. The trial stopped after treating 34 patients. The first row of this table gives the prior mean toxicity probabilities at all doses. The trial started with the lowest dose 1, as given by the algorithm. From page D-3, we see that patient 1 is not toxic. So we derive the likelihood, update the posterior distribution and calculate the posterior mean toxicity probabilities at all doses. Then for the second patient, we apply the dose whose posterior mean toxicity probability is below 0.33 but closest to 0.33. From this table, we see that dose 4 is selected. From the results on page D-3, we see that patient 2 is not toxic. We derive the likelihood again, update the posterior distribution and calculate the posterior mean toxicity probabilities at all doses. The results are given in the third row of the table. For the third patient, we select the dose with the posterior mean toxicity probability that is below 0.33 but closest to 0.33. Dose 4 is selected again, the patient is not toxic from page D-3. We derive the likelihood again, update the posterior distribution and calculate the posterior mean toxicity probabilities at all doses. The results are given in the fourth row of the table. For the fourth patient, we select the dose with the posterior mean toxicity probability that is below 0.33 but closest to 0.33. This time, after observing two non-toxic results at dose 4,

we escalate to dose 5. This seems intuitive and reasonable. After treatment, the fourth patient is still non-toxic, so we escalate to dose 6. But a toxic response was observed. Simulation showed that after 1 day, toxicity will be observed. That is, for patient 6, toxicity of patient 5 was observed. After observing a toxic response, we derive the likelihood, update the posterior distribution and calculate the posterior mean toxicity probabilities at all doses. The process was repeated and starting at patient 8, all patients were treated at dose 6. After this, patients 8, 11, 14, 15, 25, 27, and 31 are toxic. Their times to observed toxicity were 2, 0, 1, 5, 1, 1, 1 day(s). Finally, dose 6 is identified as the MTD after convergence. Of the 34 patients treated, a total of 8 patients are toxic, or 23.5%. The sample mean and variance of the toxicity probabilities of the 34 treated patients are respectively 0.227 and 0.003. Finally, at the time of stopping, the final estimated dose probabilities are 0.0068 (at dose 1), 0.0160 (at dose 2), 0.0346 (at dose 3), 0.0684 (at dose 4), 0.1231 (at dose 5), and 0.2017 (at dose 6). Dose 6 is correctly identified as the MTD, the simulated proportion of toxic patients is slightly higher than the estimated toxicity probability at dose 6, and the sample mean toxicity probability of the 34 patients in the trial is slightly higher than the identified dose probability at MTD.

For another dose trace illustrated on page D-4, the dose selection process stopped after treating 35 patients. Of the 35 patients treated, a total of 9 patients are toxic, or 25.7%. The sample mean and variance of the toxicity probabilities of the 35 treated patients are respectively 0.223 and 0.006. At the time of stopping,

the final estimated dose probabilities are 0.0061 (at dose 1), 0.0132 (at dose 2), 0.0266 (at dose 3), 0.0502 (at dose 4), 0.0882 (at dose 5), and 0.1439 (at dose 6). Dose 6 is correctly identified as the MTD, the simulated proportion of toxic patients is slightly higher than the estimated toxicity probability at dose 6, and the sample mean toxicity probability of all 35 treated patients is also slightly higher than the identified dose probability at MTD. Except for 14 patients, all other patients received dose level 6 for the treatment. The times to toxicity of the toxic patients are: 1 for patient 3, 2 for patient 8, 1 for patient 13, 3 for patient 17, 0 for patient 18, 2 for patient 20, 13 for patient 21, 7 for patient 23, and 10 for patient 24. These times were used to update the likelihood function.

For the above simulations, we used the $Beta(2, 2)$ distribution. This distribution is symmetric about 0.5 and gives equal weights to both large and small values in $(0, 1)$. We have extended this to a general beta distribution $Beta(a, b)$. When the first parameter a is smaller than b, we give more weights to small values in $(0, 1)$ than large values. This means that both the prior mean and posterior mean of the toxicity probability function $\pi_i(\alpha)$ tend to be small for all doses $i = 1, 2, \cdots, 6$. Because of this, the MTD is likely to be at dose 6. At the same time, because all toxicity probabilities are small and dose 6 is the MTD in most times, we expect relatively fast convergence.

On the other hand, we simulated with the $Beta(a, b)$ distribution when a is larger than b. In this case, we give more weights to large values in $(0, 1)$, and

therefore, most or all the prior and posterior mean toxicity probabilities are large, and in almost all cases larger than 0.33. As a result, no MTD were selected according to our criteria.

4.4 Simulation comparison with other designs with late-onset toxicity

In this section, we carry out simulation studies to compare the performance of our new design with existing designs when the toxicity is late-onset. The procedure for dealing with late-onset toxicity is the same as in the previous section. The comparison is based on the following three criteria: (1) the rate of convergence of the design, (2) the average number of toxic patients in the trial, and (3) the average toxicity of all patients treated in the trial. All simulations are based on 1000 replications. All simulations use the $Beta(a, b)$ prior where $a = b = 2$ which are very good as we have observed in Section 4.3. The stopping rule is the same for all designs. That is, we stop each design when the absolute difference of the toxicity probabilities of the last two chosen doses is less than 0.005. The dose selection rule is also the same. That is, for each patient, the dose selected is the one whose toxicity probability is less than 0.33 but at the same time closest to 0.33.

The motivation of using these three criteria is as follows. The rate of convergence tells us how many patients we treat, on average, before we stop the

trial. The intuition is that the less patients we have in the trial, the less number of patients are subject to toxicity. For a fixed number of patients in the trial, the less number of toxic patients, the better the design. So we wish to estimate the average number of toxic patients for each design. Finally, we estimate the average toxicity probability of all patients in the trial. We hope that the lower this average, the better the design, because this average tells us the overall level of toxicity of all patients treated in the trial.

For our new design, the dose toxicity probability function is given by $\pi_i(\alpha) = \frac{2\Phi(-3+\alpha d_i,\mu,\sigma^2)}{1+\Phi(-3+\alpha d_i,\mu,\sigma^2)}$, where α is the unknown parameter following the $Beta(2,2)$ prior distribution, μ is the mean of the normal distribution and σ^2 is the variance of the normal distribution, and $d_i = i, i = 1, 2, \cdots, 6$, is the dose. The dose toxicity probability function of the logistic design is given by $\pi_i(\alpha) = \frac{exp(-3+\alpha d_i)}{1+exp(-3+\alpha d_i)}$, where α is the unknown parameter following the $Beta(2,2)$ prior distribution and $d_i = i, i = 1, 2, \cdots, 6$, is the dose. The dose toxicity probability function of the hyperbolic tangent design design is given by $\pi_i(\alpha) = \left[\frac{tanh(-3+d_i)+2}{2}\right]^{\alpha}$, where α is the unknown parameter following the $Beta(2,2)$ prior distribution and $d_i = i, i = 1, 2, \cdots, 6$, is the dose.

Simulation results are summarized in Appendix E. For our new design, we simulated the cases of $N(0, 0.5)$, $N(0, 1)$, and $N(0, 2)$. From Table E.1 on page E-2, we see our new design with $N(0, 1)$ performs very good. The average stopping times for our new design (17.672 for $\sigma^2 = 0.5$, 24.85 for $\sigma^2 = 1$, and 21.286 for

$\sigma^2 = 2$) are reasonably good compared with the logistic (23.064) and hyperbolic tangent (19.078) designs, so our new design is efficient, although the hyperbolic tangent design is also very good. This means that on average, small sample sizes are needed to complete Phase I clinical trials with late-onset toxicity using our new design. This is both logistically required and ethically desirable. Our new design with $\sigma^2 = 1$ is easy to apply because it is the standard normal distribution. For this design, the average proportion of toxic patients in the trial is 18.4%, which is lower than the average proportion of 23.1% for the logistic design and 23.7% for the hyperbolic tangent design. This implies that on average, we subject less patients in the trial to toxicity if we use our new design. This means that our new design is more ethical than both logistic and hyperbolic tangent designs. Finally the average toxicity probability for all patients is 18.3% for our new design with $N(0, 1)$, that is lower than 21.8% for the logistic design and 23.5% for the hyperbolic tangent design. This also shows that our new design is more ethical.

For comparison, we have also included the histograms of MTD and the stopping time, for all three designs: new, logistic and hyperbolic tangent. See pages E-3 to page E-8.

4.5 Summary

This chapter is exclusively focused on simulation studies of the new design with late-onset toxicity and comparison with the existing parametric designs with

late-onset toxicity. Simulation results confirm the advantage of our new design because we can adjust the variance of the normal distribution in order to model different possible patterns of the dose-toxicity relationship. Simulation studies show that the new design performs much better than the three existing designs when the toxicity is late-onset and follows the geometric distribution. Different criteria are used, including the convergence rate and average proportion of toxic patients. Finally the MTD is identified together with its toxicity probability.

Chapter 5

Conclusion

5.1 Summary of achievements

In this MSc thesis, I have introduced a new parametric design of Phase I clinical trials and compared its performance with existing parametric designs in the literature. All the designs considered use the continual reassessment method (CRM).

The new design was introduced after I recognized that instead of the exponential function in the logistic design, the cumulative distribution function has a similar shape and may be used after some modifications. The new design of the dose-toxicity relationship is an increasing function and has properties similar to the existing parametric designs of Phase I clinical trials.

We not only assess the convergence and properties of the new design, but also compare its performance with the existing parametric designs with regard to three important criteria: (1) the identification of the maximum tolerated dose (MTD),

(2) the convergence to the MTD, and (3) the proportion of toxic patients. Based on these criteria, simulation results have demonstrated our new design is both ethically and statistically better than the parametric designs in the literature, whether the toxicity is immediately observed or late-onset.

5.2 Future research

Design of Phase I clinical trials is very important in biostatistics because it forms the foundation for all future phases of clinical trials. The goal of Phase I clinical trials is to identify the maximum tolerated dose (MTD) to be recommended for further study in Phase II. If the MTD is incorrectly identified, substantial harms may be done. If the MTD is under-estimated, then the drug is potentially useless because its dose is ineffective. On the other hand, if the MTD is over-estimated, then it will be harmful and overly toxic, and many patients may die.

The design of Phase I clinical trials depends on many factors. Therefore the design can be improve if practical considerations are included. For example, important patient's covariates such as weights, age and gender may be included. In this case, regression models may be required.

Bibliography

Babb, J., Rogatko, A. and Zacks, S. (1998). Cancer phase i clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine*, **17** 1103–1120.

Cheung, Y. (2011). *Dose finding by the continual reassessment method*. CRC Press, New York, USA.

Durham, S., Flournoy, N. and Rosenberger, W. (1997). A random walk rule for phase i clinical trials. *Biometrics*, **53** 745–760.

Hedaya, M. A. (2007). *Basic Pharmacokinetics*. CRC Press, Boca Raton, FL.

Hoeting, J., Madigan, D., Raftery, A. E. and Volinsky, C. (1999). Bayesian model averaging: a tutorial. *Statistical Science*, **14** 382–401.

Jeffreys, H. (1961). *Theory of Probability*. 3rd ed. Oxford University Press.

Kummar, S., Rubinstein, L., Kinders, R., Parchment, R., Gutierrez, M., Murgo, A., Ji, J., Mroczkowski, B., Pickeral, O., Simpson, M., Hollingshead, M., Yang, S., Helman, L., Wiltrout, R., Collins,

J., Tomaszewski, J. and Doroshow, J. (2008). Phase 0 clinical trials: Conceptions and msconceptions. *The Cancer Journal*, **14** 133–137.

Lin, Y. and Shih, W. (2001). Statistical properties of the traditional alogrithm-based designs for phase i cancer clinical trials. *Biostatistics*, **2** 203–215.

MDAnderson (Accessed 2016-06-28). Clinical trials at md anderson: Is a clinical trial the right treatment choice for you? URL https://www.mdanderson.org/patient-and-cancer-information/cancer-information/clinical-trials/clinicaltrials-1111.pdf.

O'Quigley, J. and Chevret, S. (1991). Methods for dose finding studies in cancer clinical trials: A review and results of a monte carlo study. *Statistics in Medicine*, **10** 1647–1664.

O'Quigley, J., Pepe, M. and Fisher, L. (1990). Continual reassessment method: a practical design for phase i clinical trials in cancer. *Biometrics*, **46** 33–48.

O'Quigley, J. and Shen, L. (1996). Continual reassessment method: a likelihood approach. *Biometrics*, **52** 673–684.

Rosenberger, W. and Lachin, J. M. (2002). *Randomization in Clinical Trials: Theory and Practice*. John Wiley & Sons, New York.

Simon, R., Freidlin, B., Rubinstein, L., Arbuck, S., Collins, J. and

Christian, M. (1997). Accelerated titration designs for phase i clinical trials in oncology. *Journal of the National Cancer Institute*, **89** 1138–1147.

Storer, B. (1989). Design and analysis of phase i clinical trials. *Biometrics*, **45** 925–937.

Stylianou, M. and Flournoy, N. (2002). Dose finding using the biased coin up-and-down design and isotonic regression. *Biometrics*, **58** 171–177.

UnitedNations (Accessed 2016-06-28). URL http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf.

Yin, G. (2012). *Clinical trial design: Bayesian and frequentist adaptive methods*. Wiley.

Yin, G. and Yuan, Y. (2009). Bayesian model averaging continual reassrssment method in phase i clinical trials. *Journal of the American Statistical Association*, **104** 954–968.

Yuan, Y. and Yin, G. (2011). Bayesian hybrid dose-finding design in phase i oncology clinical trials. *Statistics in Medicine*, **30** 2098–2108.

Appendix A

List of terms and symbols

- MTD: maximum tolerated dose
- DLT: dose limiting toxicity
- FDA: Food and Drug Administration (USA)
- ATD: accelerated titration design
- BCD: biased coin dose-finding method
- CRM: continual reassessment method
- BMA: Bayeasian model averaging
- EWOC: escalation with overdose control
- $\Phi(x, \mu, \sigma^2)$: cumulative distribution function of the normal distribution $N(\mu, \sigma^2)$

- $d_i, i = 1, 2, \cdots, K$: dose levels
- $p_i, i = 1, 2, \cdots, K$: prespecified toxicity probability at dose level d_i
- ϕ_T: target toxicity probability
- Y_j: binary toxicity outcome (1 =toxicity, 0 =non-toxicity)
- π_i: probability of toxicity at dose level d_i under the dose-toxicity model
- $L(D|\alpha)$: likelihood function

Appendix B

Graphs - immediate toxicity

Figure B.1: Graph of function $f(x) = \frac{2\Phi(x,0,1)}{1+\Phi(x,0,1)}$ with $\mu = 0$ and $\sigma^2 = 1$

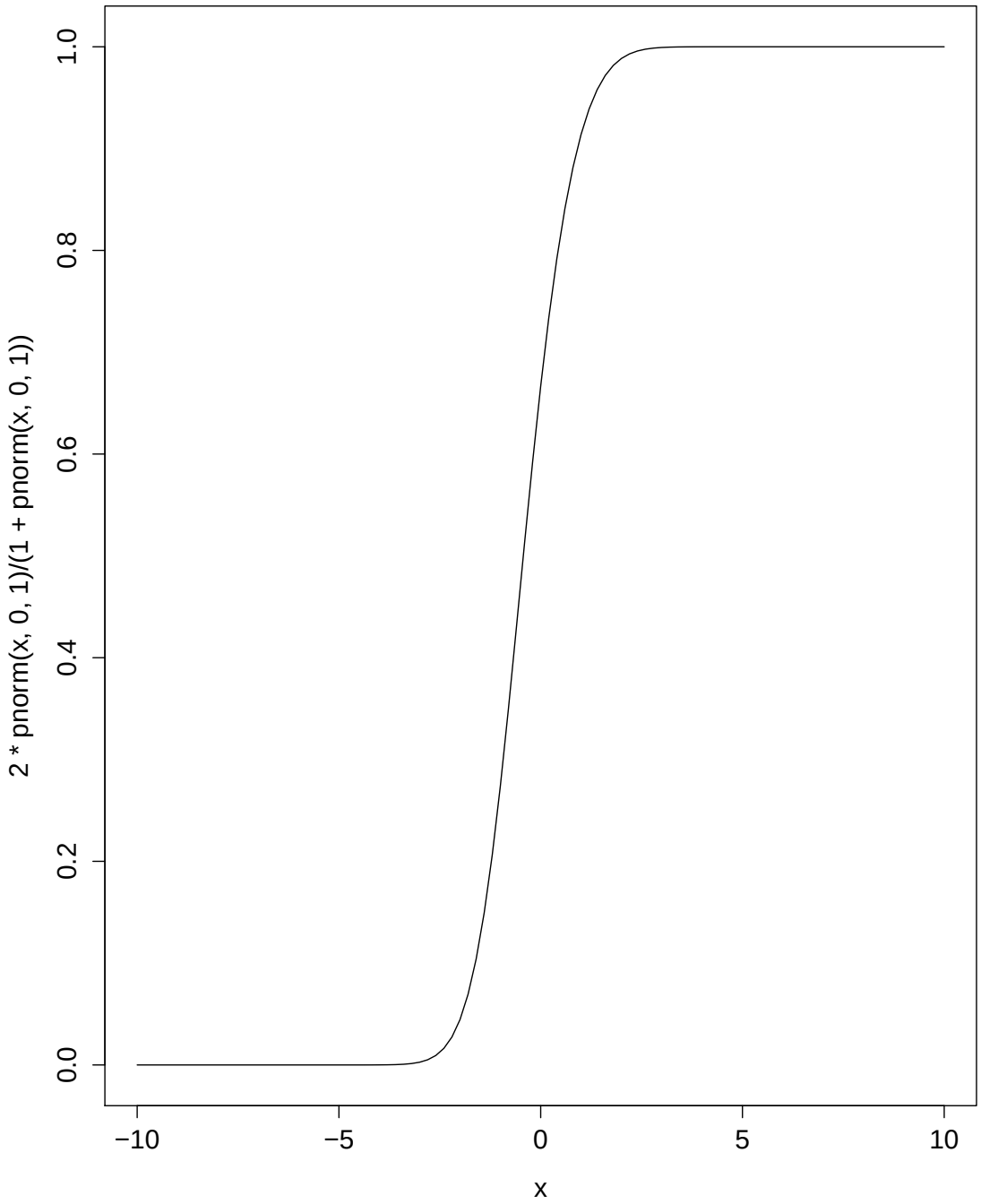


Figure B.2: Graph of function $f(x) = \frac{2\Phi(x,0,2)}{1+\Phi(x,0,2)}$ with $\mu = 0$ and $\sigma^2 = 2$

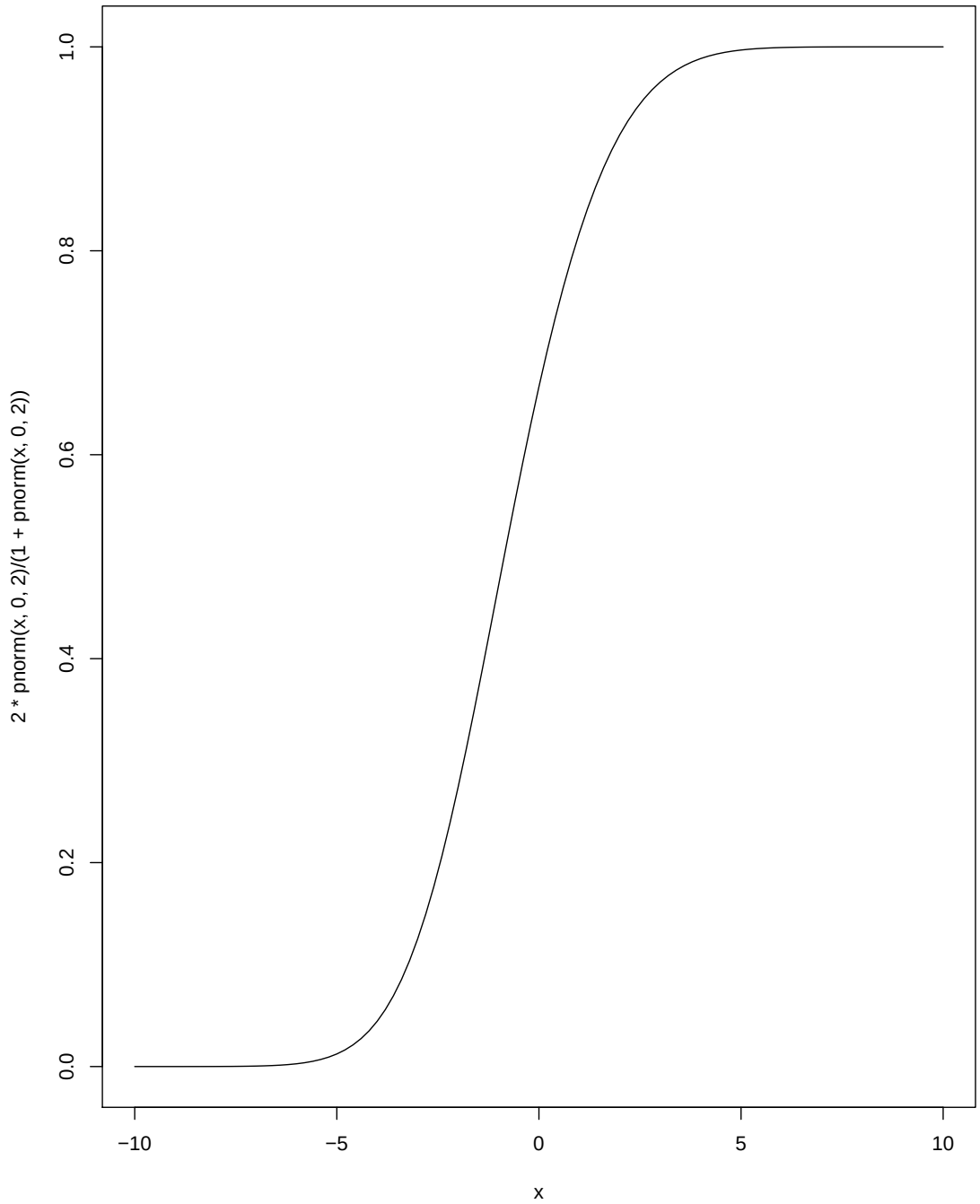


Figure B.3: Graph of function $f(x) = \frac{2\Phi(x,0,3)}{1+\Phi(x,0,3)}$ with $\mu = 0$ and $\sigma^2 = 3$

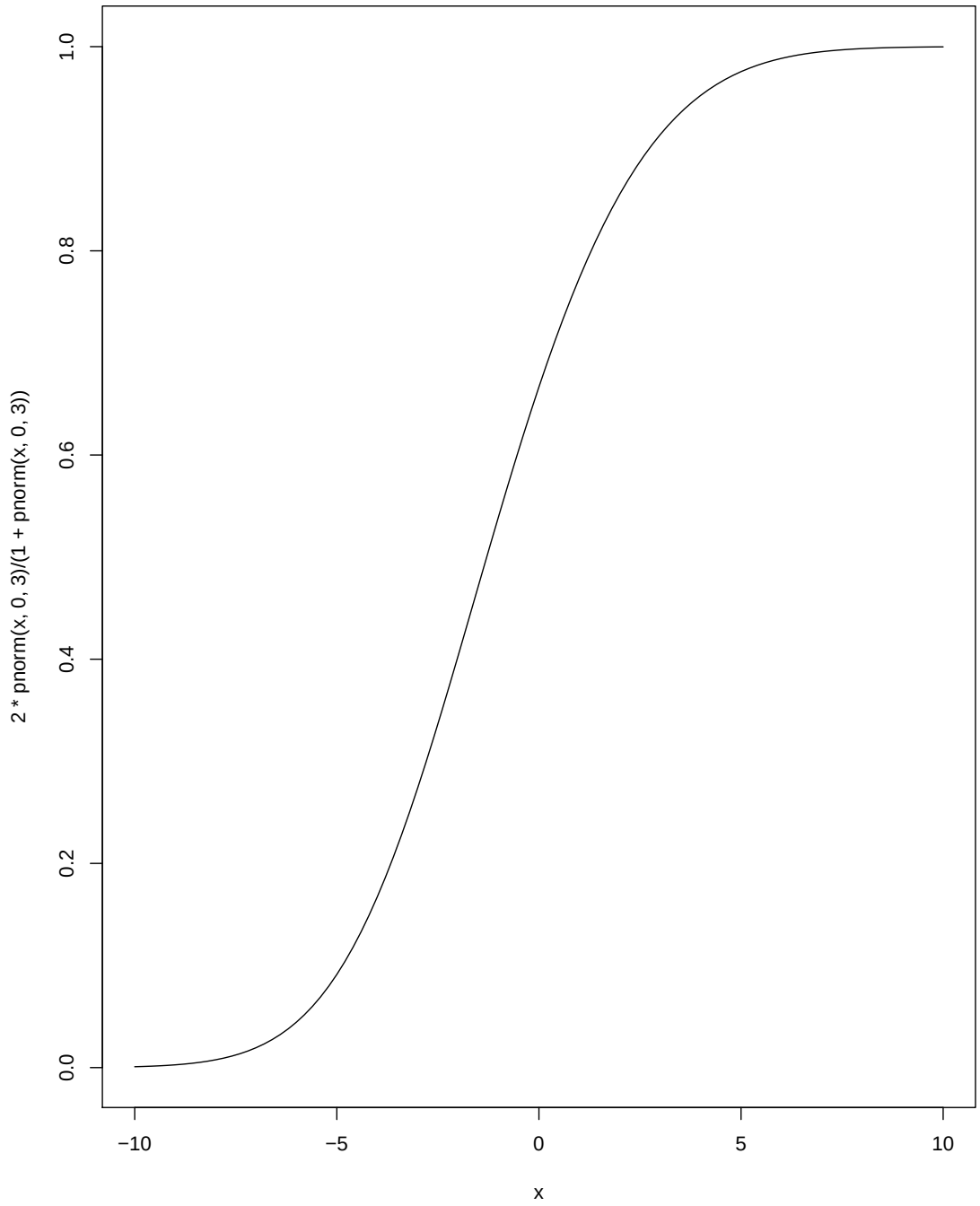


Figure B.4: Graph of function $f(x) = \frac{2\Phi(x,0,4)}{1+\Phi(x,0,4)}$ with $\mu = 0$ and $\sigma^2 = 4$

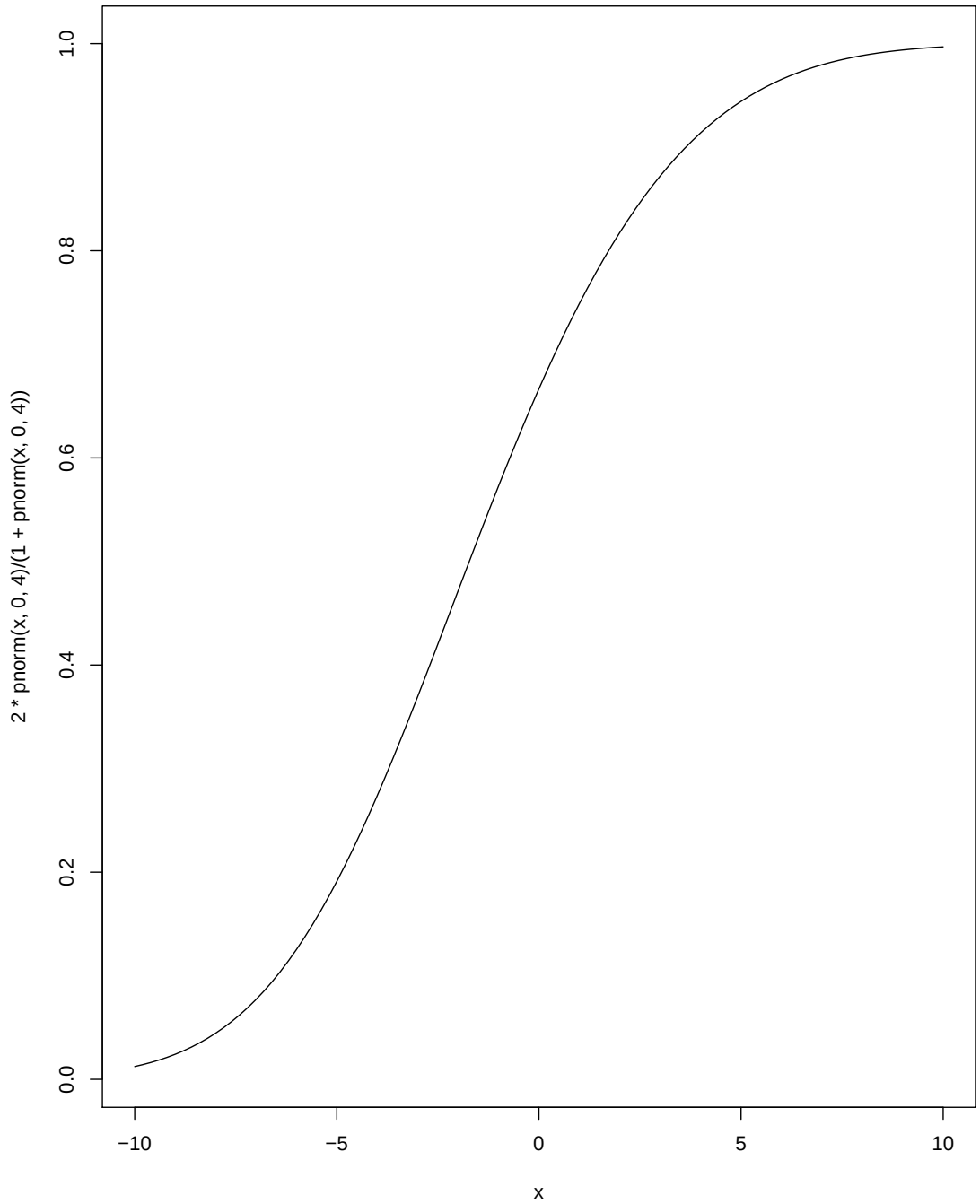


Figure B.5: Graph of function $f(x) = \frac{2\Phi(x,0,5)}{1+\Phi(x,0,5)}$ with $\mu = 0$ and $\sigma^2 = 5$

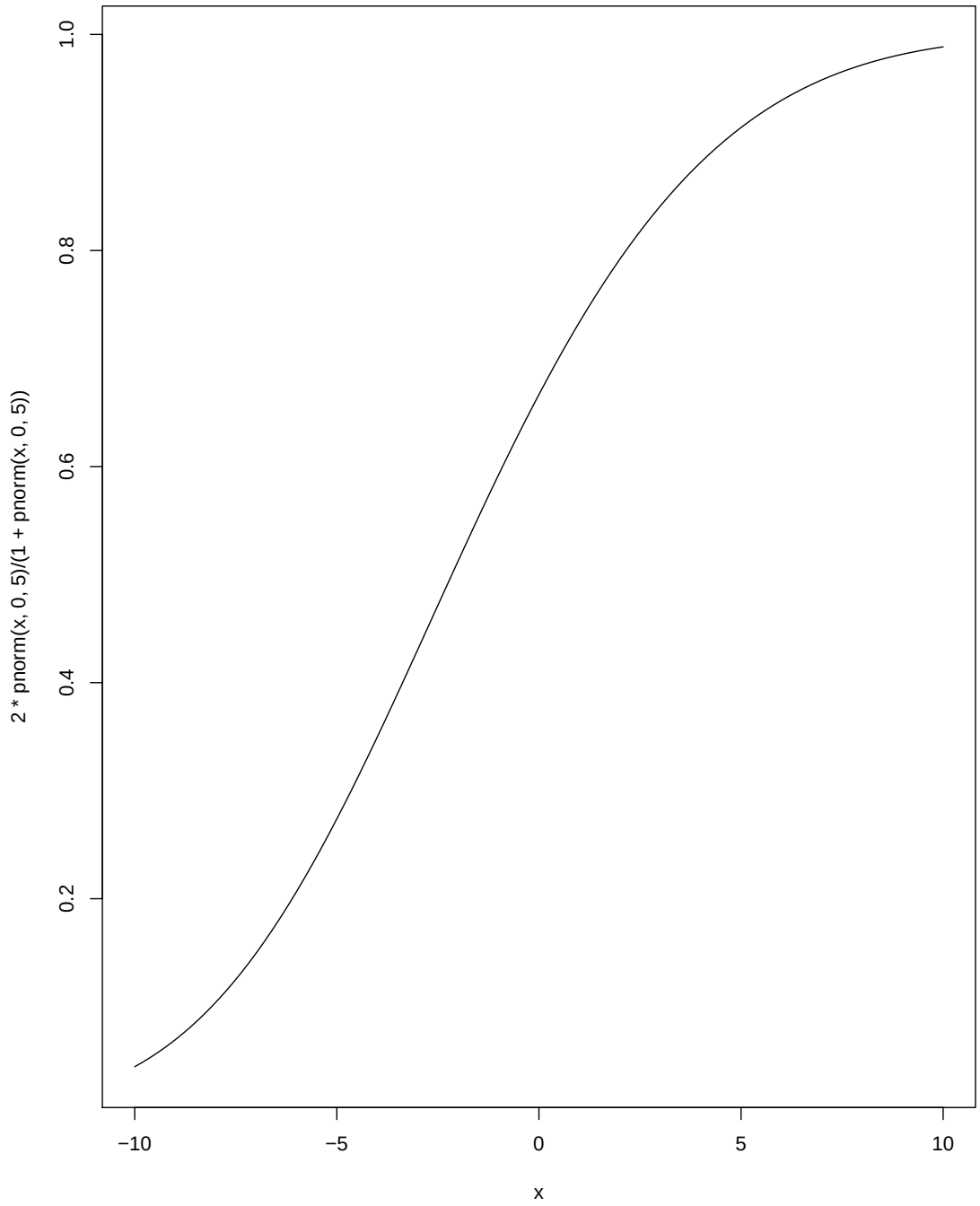


Figure B.6: Graph of function $f(x) = x^{exp(-2)}$

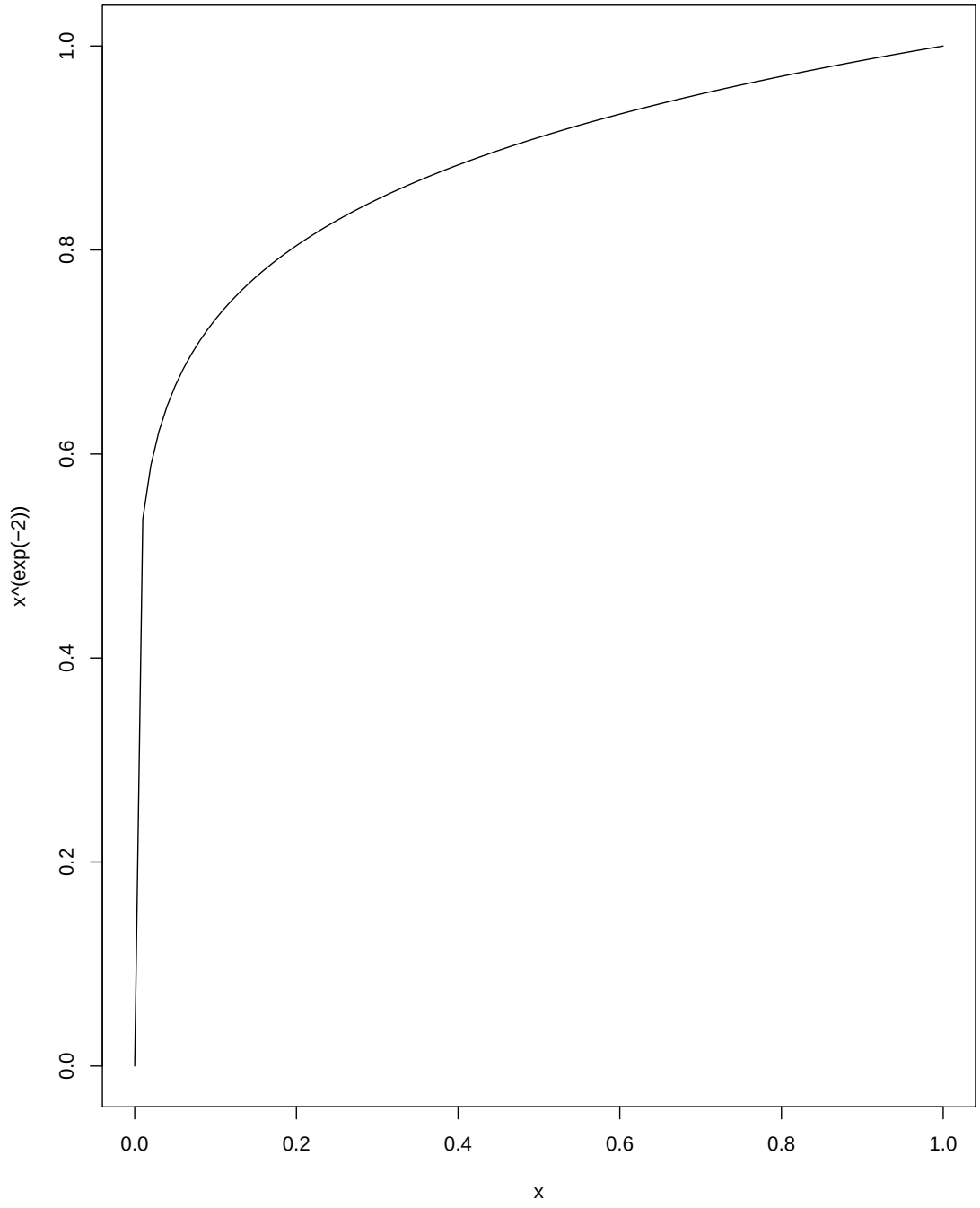


Figure B.7: Graph of function $f(x) = x^{exp(-1)}$

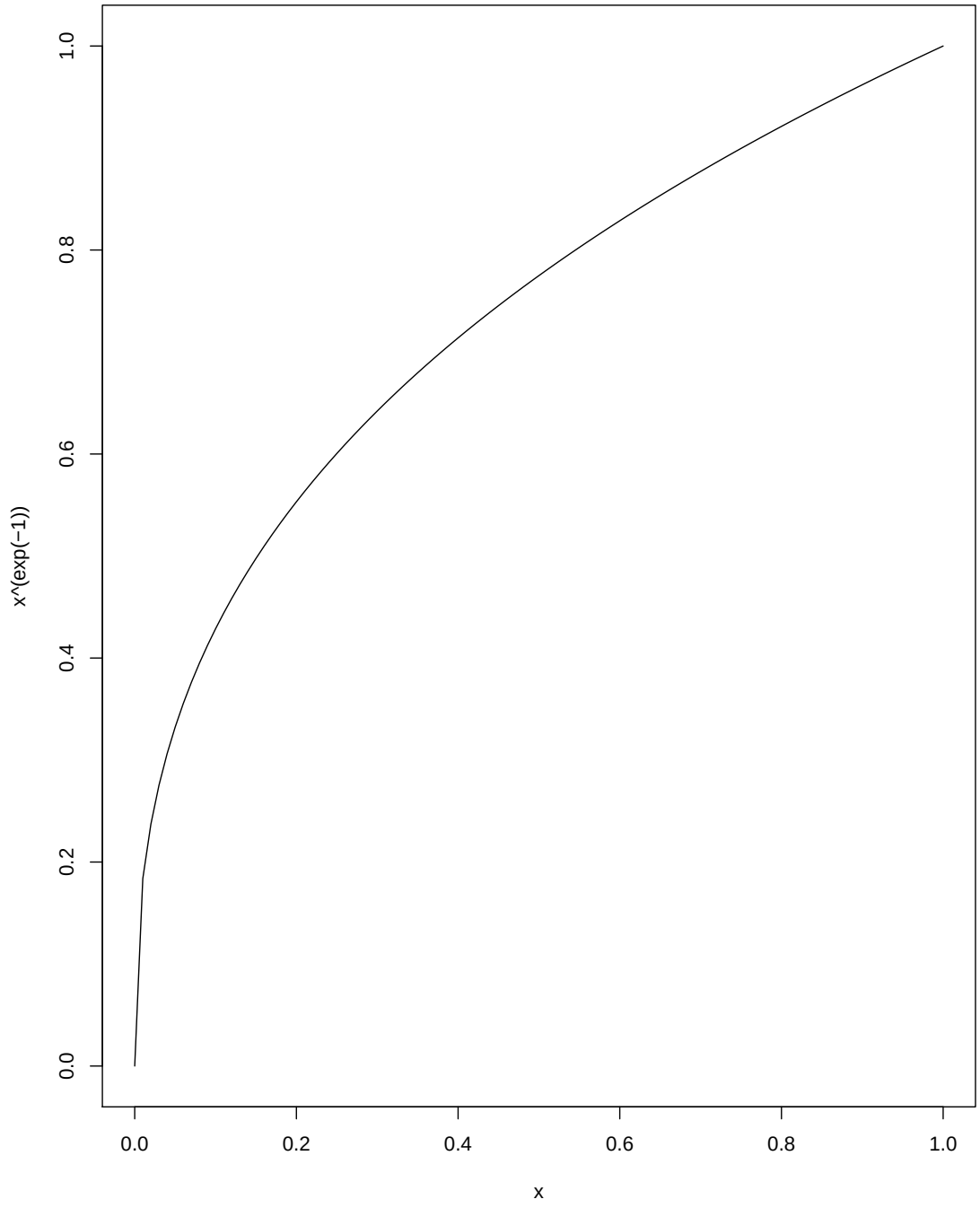


Figure B.8: Graph of function $f(x) = x^{exp(0)}$

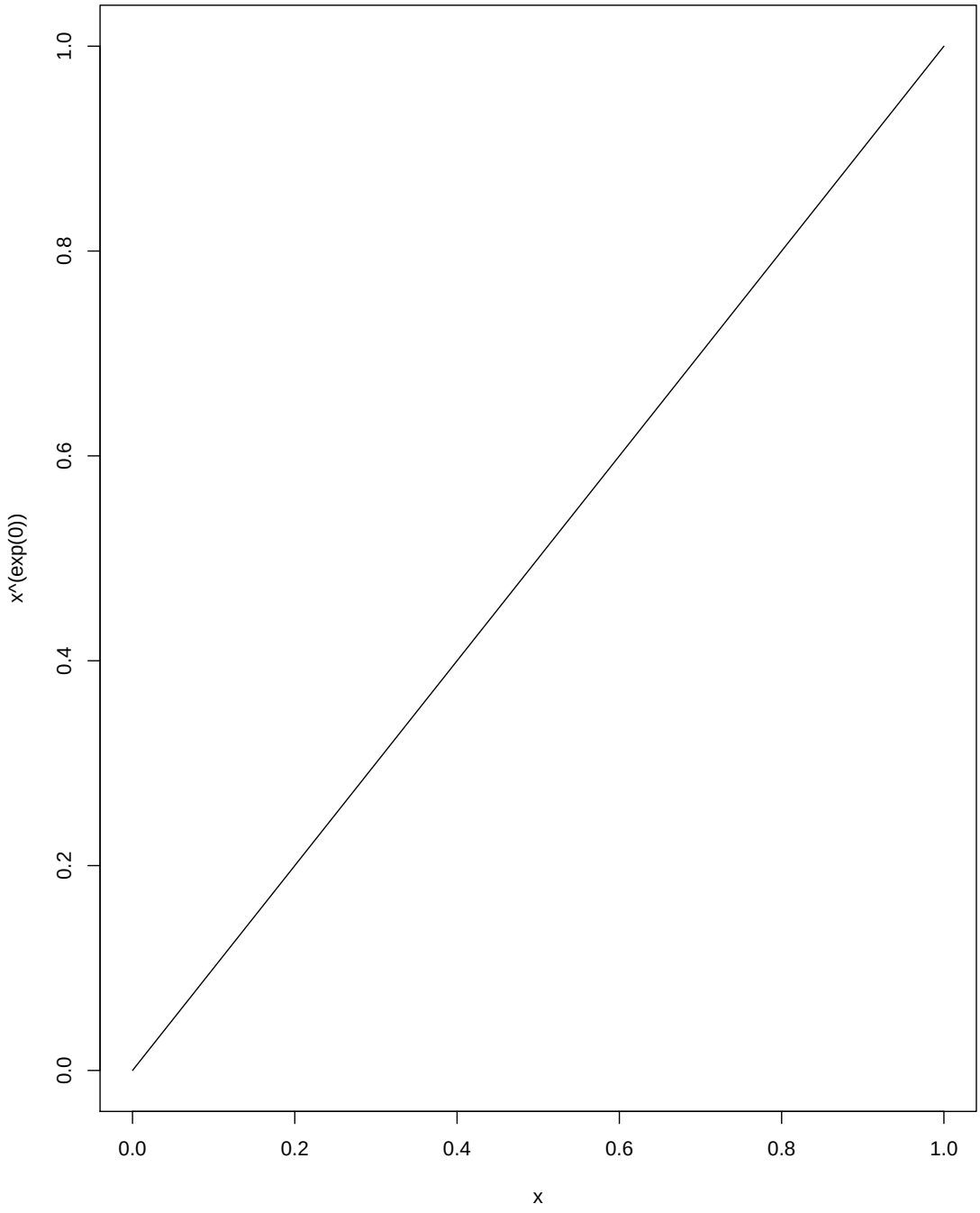


Figure B.9: Graph of function $f(x) = x^{exp(1)}$

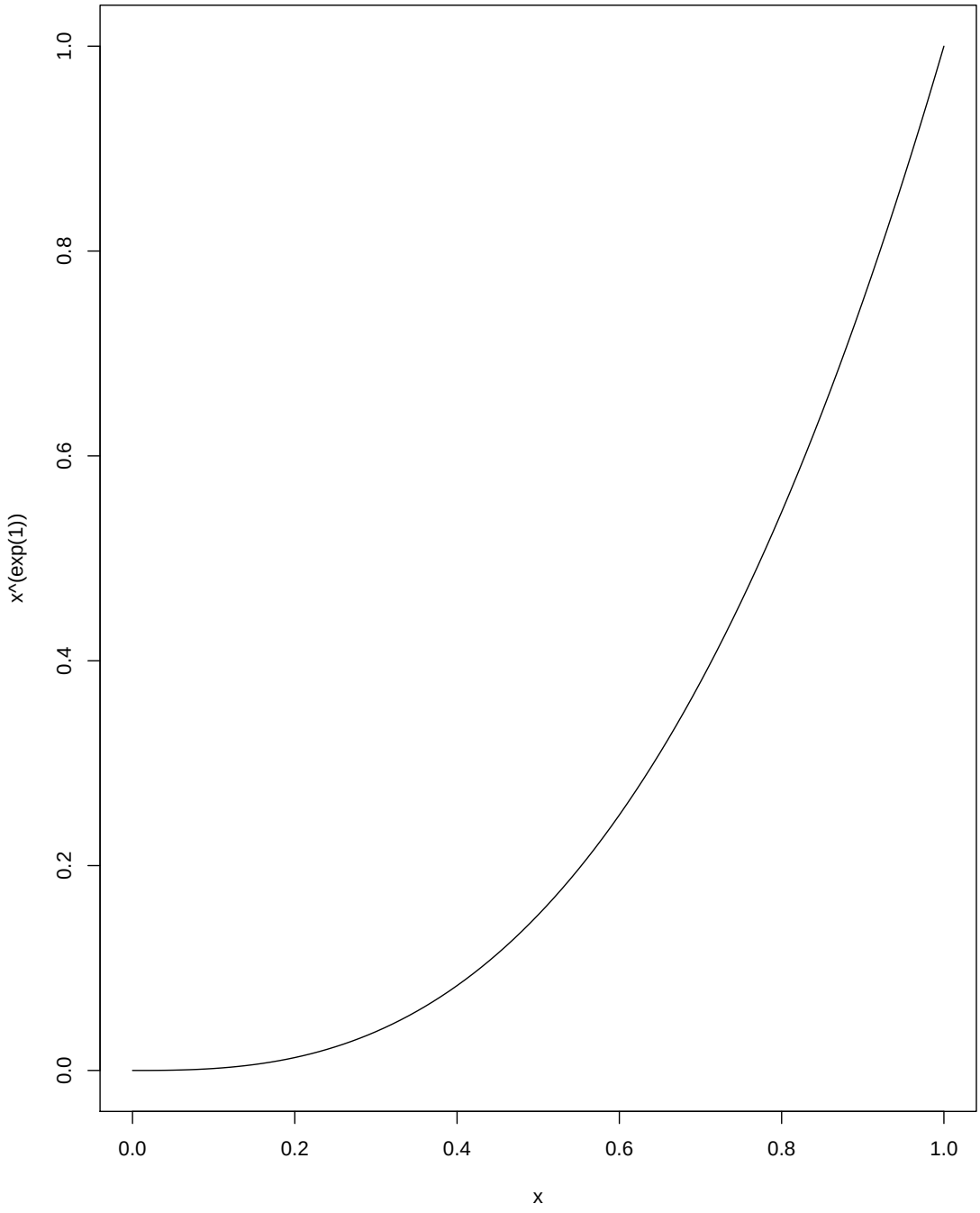


Figure B.10: Graph of function $f(x) = x^{exp(2)}$

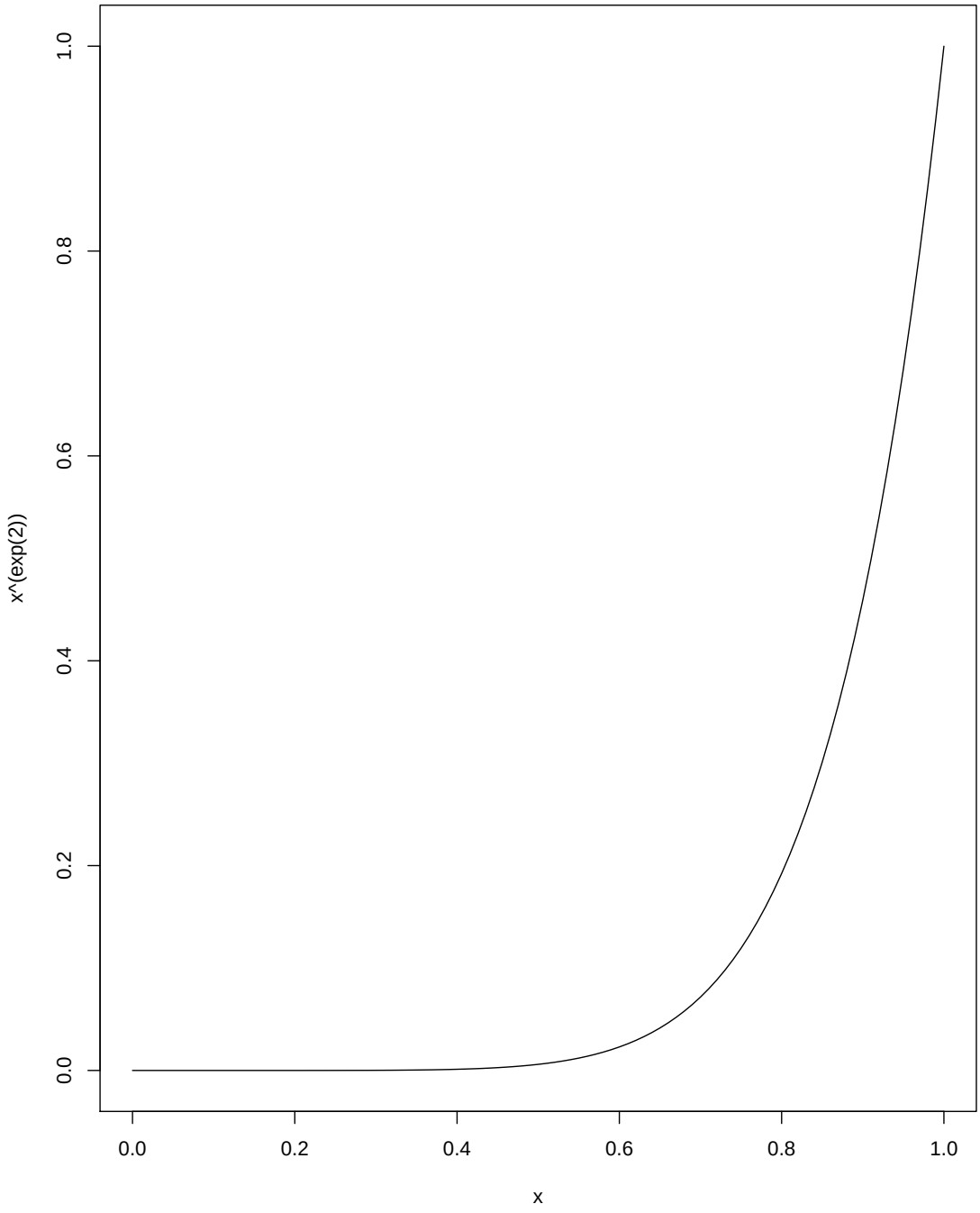


Figure B.11: Graph of function $f(x) = \frac{exp(x)}{1+exp(x)}$

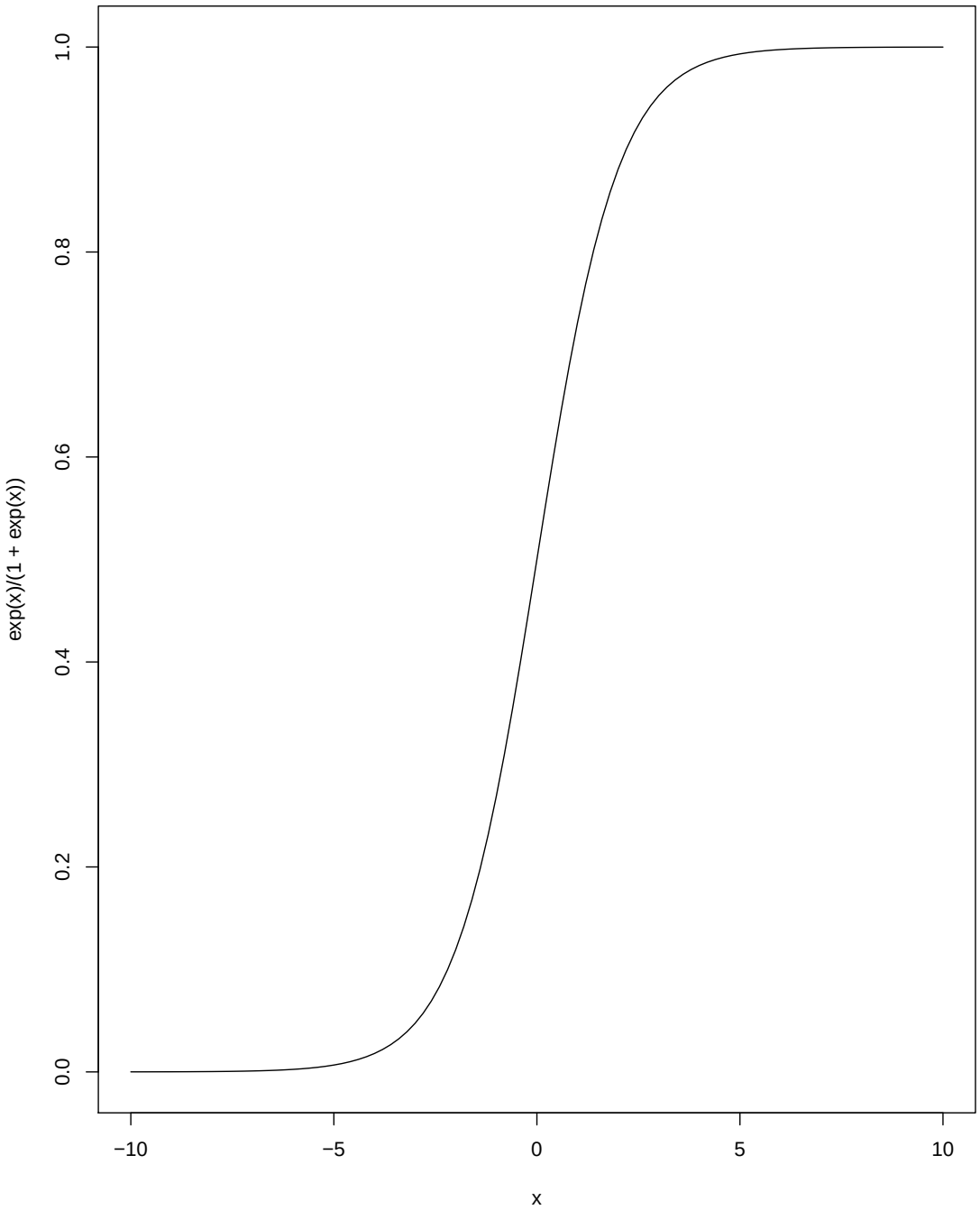


Figure B.12: Graph of function $f(x) = \left(\frac{\tanh(x)+1}{2}\right)^{0.2}$

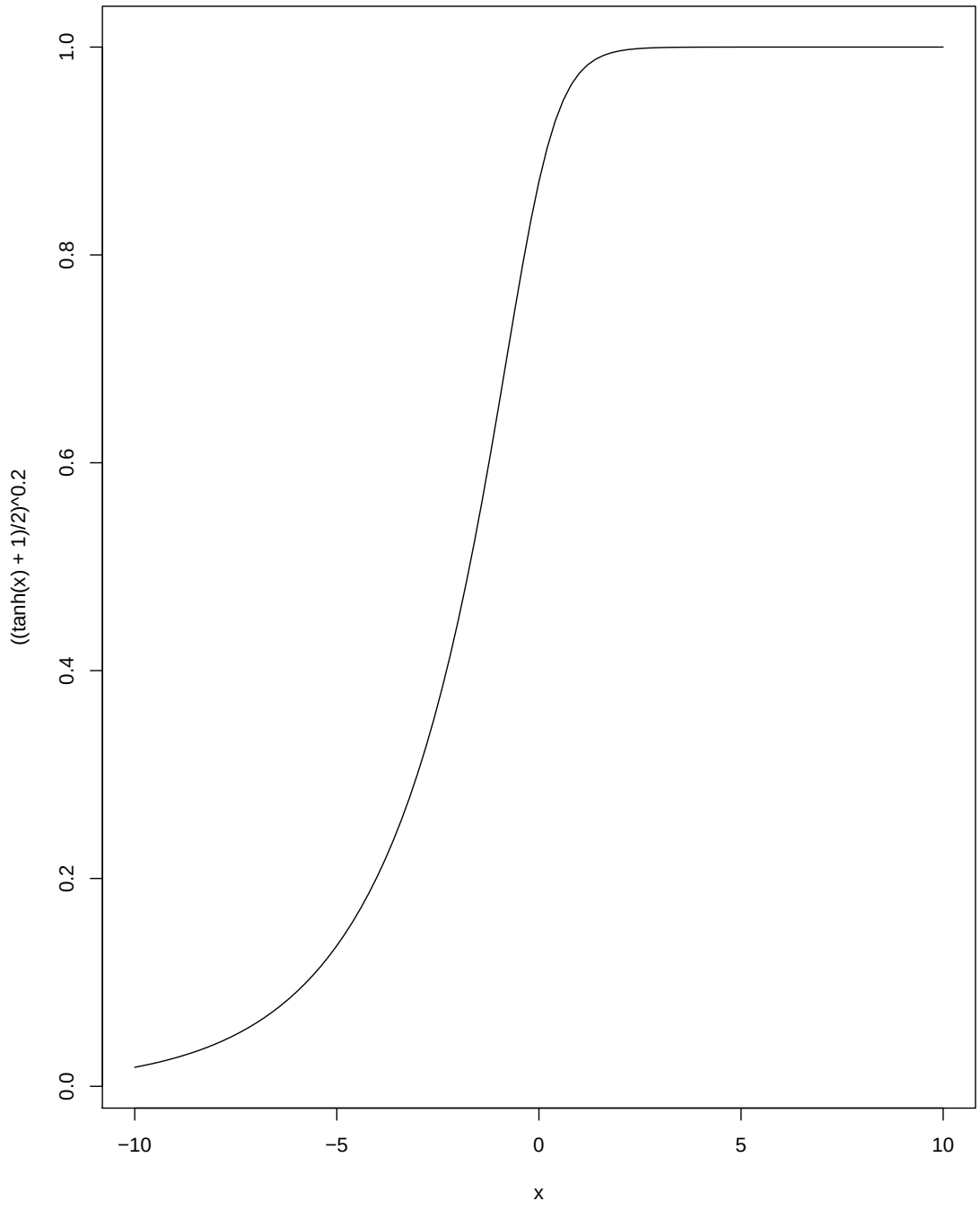


Figure B.13: Graph of function $f(x) = \left(\frac{\tanh(x)+1}{2}\right)^{0.5}$

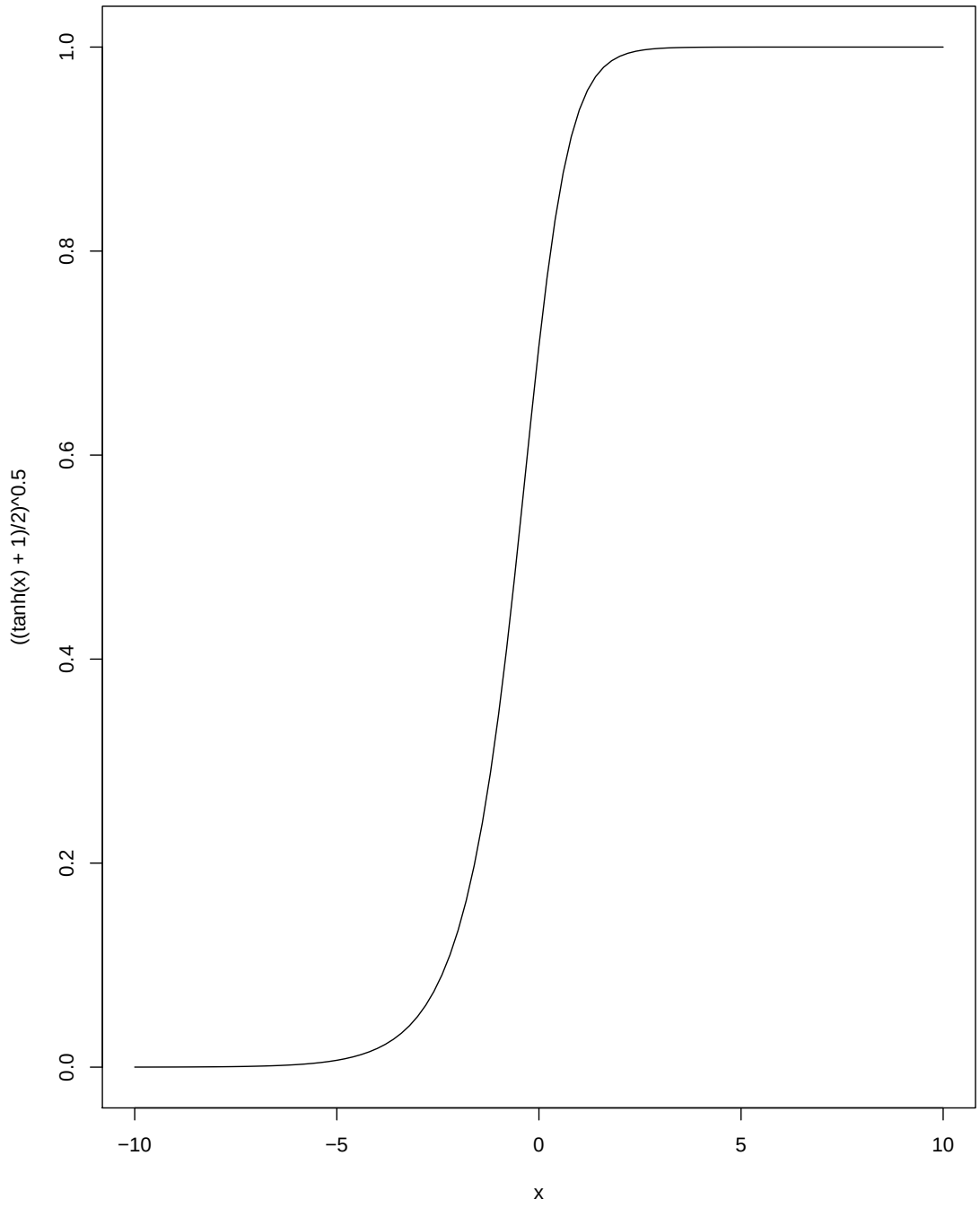


Figure B.14: Graph of function $f(x) = \frac{\tanh(x)+1}{2}$

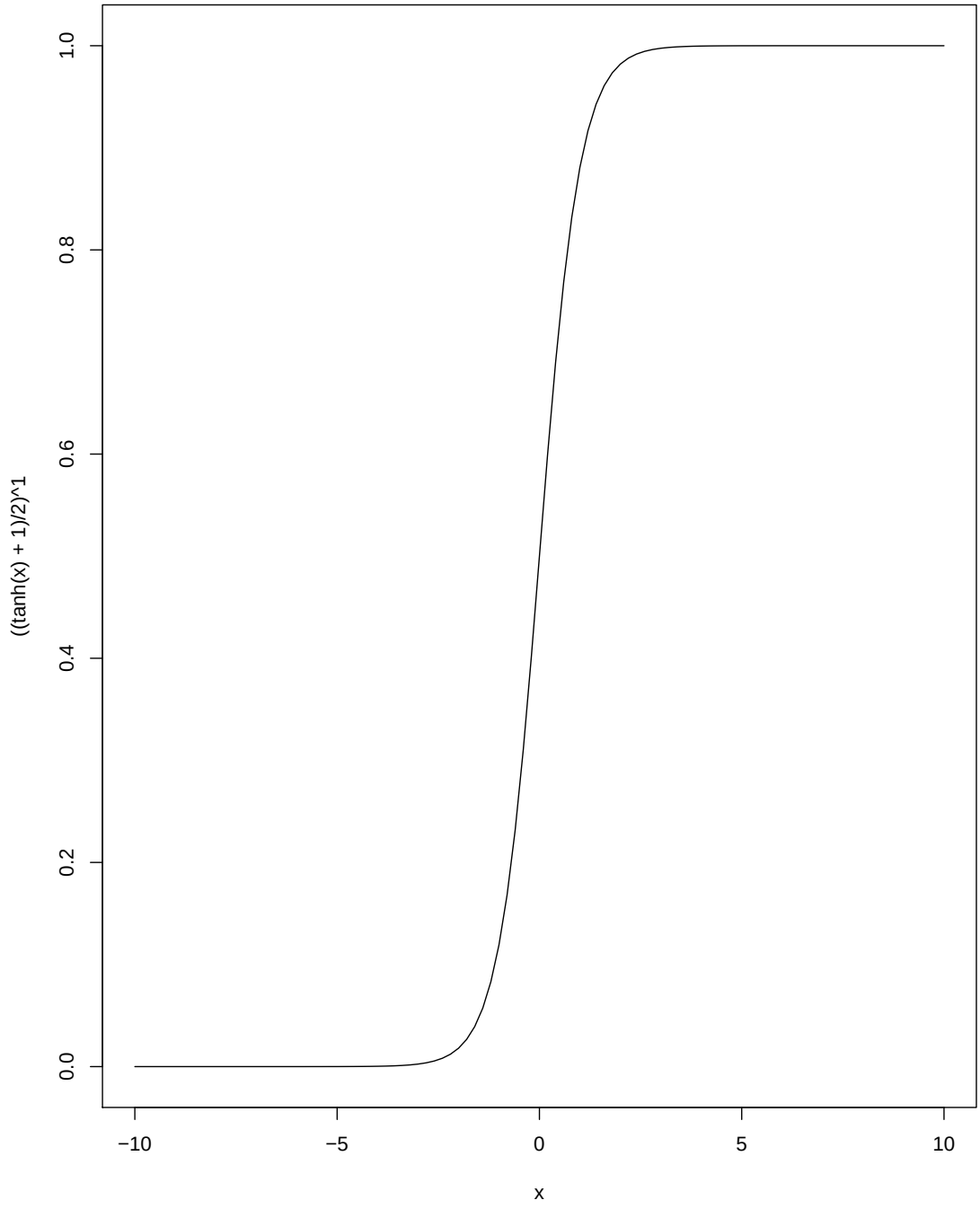


Figure B.15: Graph of function $f(x) = \left(\frac{\tanh(x)+1}{2}\right)^2$

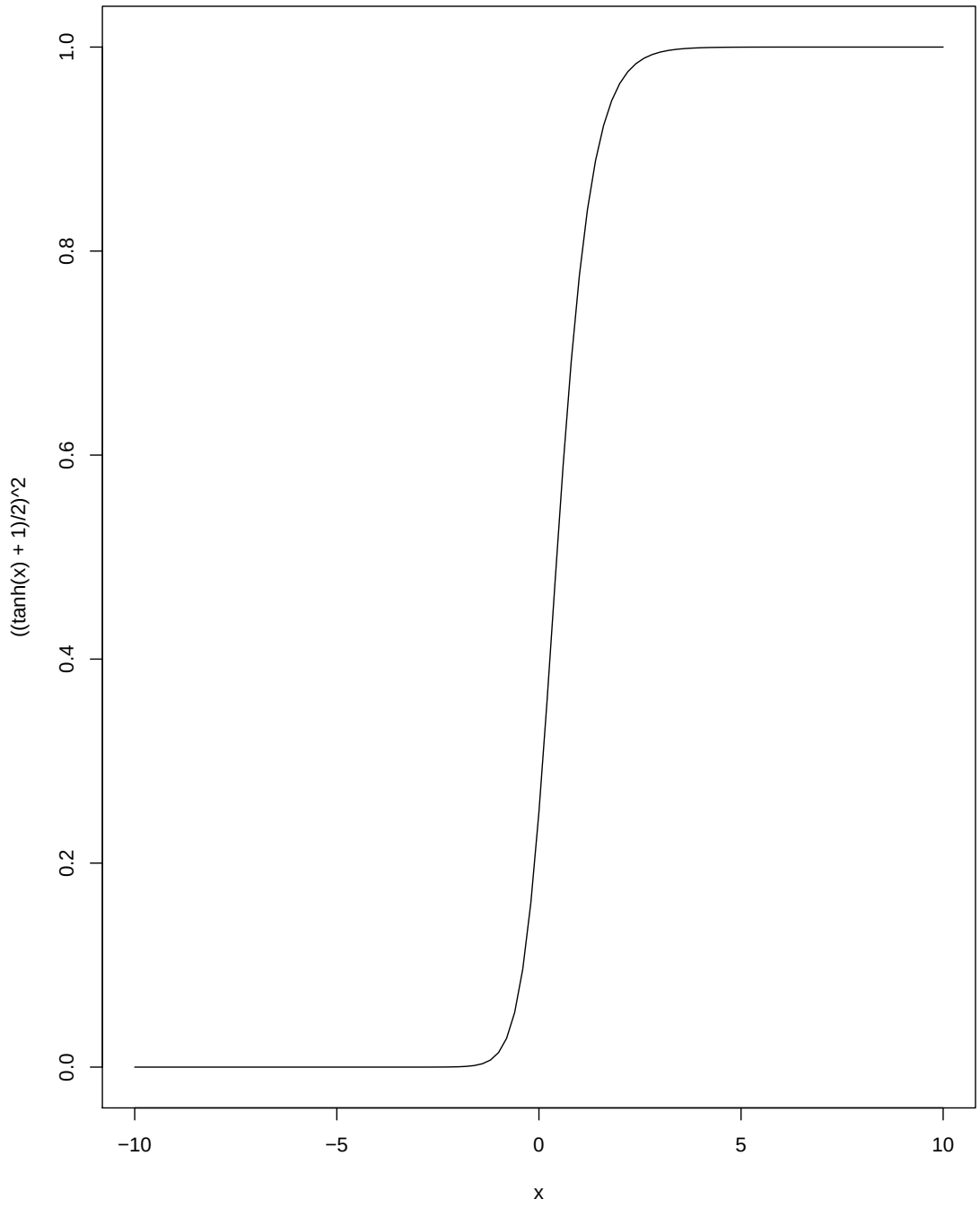


Figure B.16: Graph of function $f(x) = \left(\frac{\tanh(x)+1}{2}\right)^3$

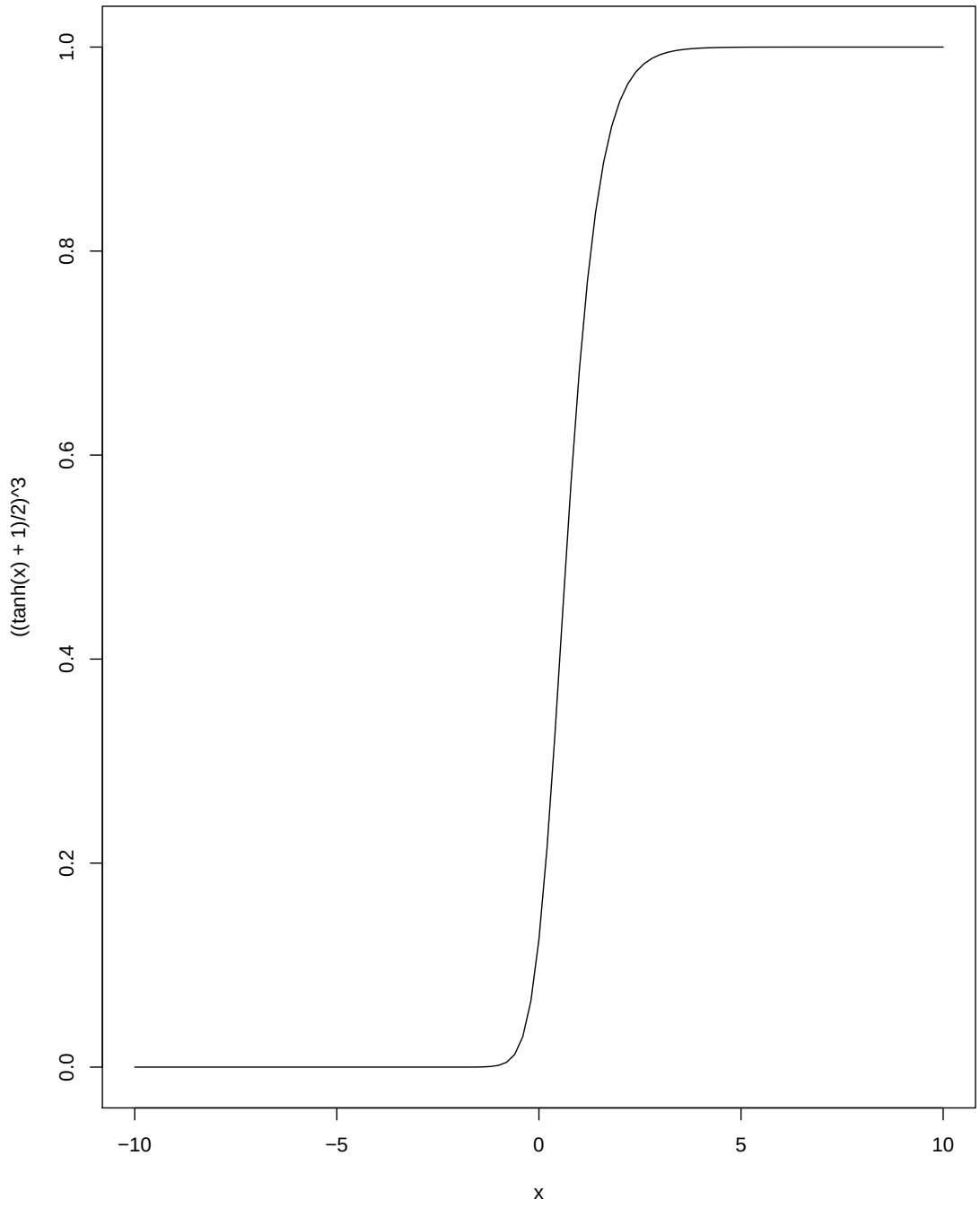


Table B.1: Boldface indicates toxicity, the italicized indicated no toxicity, and the underlined indicates convergence. The MTD is dose 3.

	Mean model toxicity probability at dose.					
Patient	1	2	3	4	5	6
1	*0.0145*	0.0638	0.1786	0.3315	0.4747	0.5880
2	0.0145	0.0633	0.1771	*0.3292*	0.4720	0.5853
3	0.0111	0.0416	0.1145	*0.2251*	0.3470	0.4590
4	0.0093	0.0308	0.0821	0.1664	**0.2696**	0.3744
5	0.0138	0.0559	0.1592	**0.3188**	0.4918	0.6412
6	0.0172	0.0777	*0.2254*	0.4362	0.6375	0.7862
7	0.0163	0.0713	*0.2070*	0.4074	0.6073	0.7614
8	0.0155	0.0660	*0.1913*	0.3820	0.5796	0.7378
9	0.0148	0.0615	*0.1779*	0.3596	0.5542	0.7153
10	0.0142	0.0577	*0.1663*	0.3397	0.5309	0.6941
11	0.0137	0.0545	0.1563	**0.3219**	0.5096	0.6741
12	0.0158	0.0679	*0.1981*	0.3988	0.6067	0.7695
13	0.0153	0.0644	*0.1876*	0.3812	0.5870	0.7528
14	0.0148	0.0613	*0.1782*	0.3651	0.5686	0.7367
15	0.0144	0.0586	*0.1698*	0.3503	0.5513	0.7212
16	0.0140	0.0562	*0.1622*	0.3368	0.5350	0.7063
17	0.0137	0.0540	0.1554	**0.3243**	0.5197	0.6919
18	0.0152	0.0635	*0.1855*	0.3809	0.5918	0.7620
19	0.0149	0.0612	*0.1784*	0.3684	0.5774	0.7495
20	0.0145	0.0591	*0.1718*	0.3568	0.5636	0.7373
21	0.0142	0.0572	*0.1658*	0.3458	0.5506	0.7255
22	0.0140	0.0554	*0.1602*	0.3356	0.5380	0.7140
23	0.0137	0.0538	0.1550	**0.3260**	0.5261	0.7028
24	0.0149	0.0612	*0.1785*	0.3707	0.5832	0.7580
25	0.0146	0.0595	*0.1732*	0.3611	0.5719	0.7480
26	0.0144	0.0579	**0.1681**	0.3519	0.5610	0.7382
27	0.0156	0.0652	*0.1915*	0.3948	0.6130	0.7855
28	0.0153	0.0635	*0.1861*	0.3854	0.6024	0.7765
29	0.0150	0.0619	*0.1811*	0.3765	0.5921	0.7677
30	0.0148	0.0604	*0.1763*	0.3680	0.5821	0.7591

Figure B.17: Illustration of Dose Selection and Toxicities with 6 Dose Levels - Immediate Toxicity

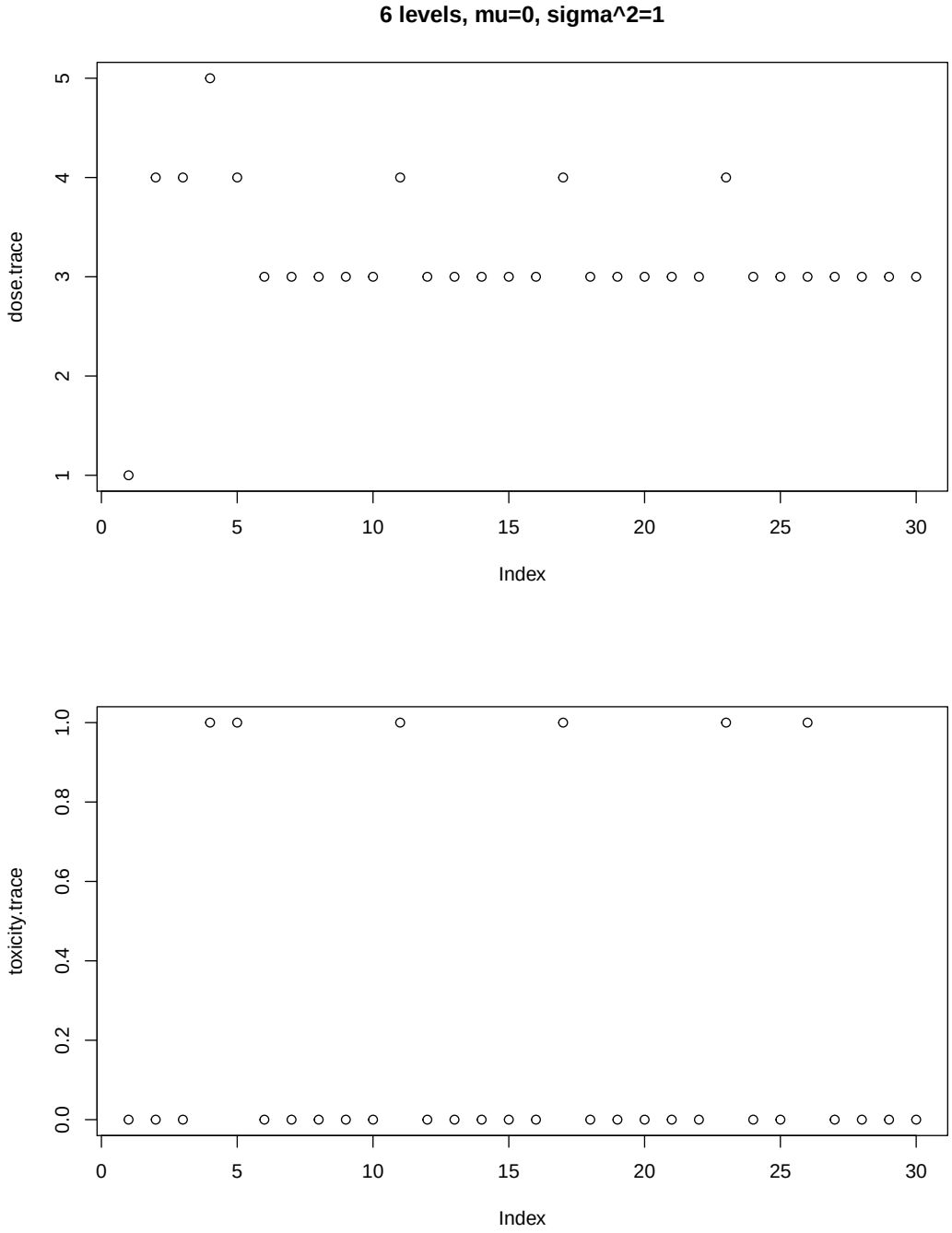


Figure B.18: Illustration of Dose Selection and Toxicities with 6 Dose Levels - Immediate Toxicity

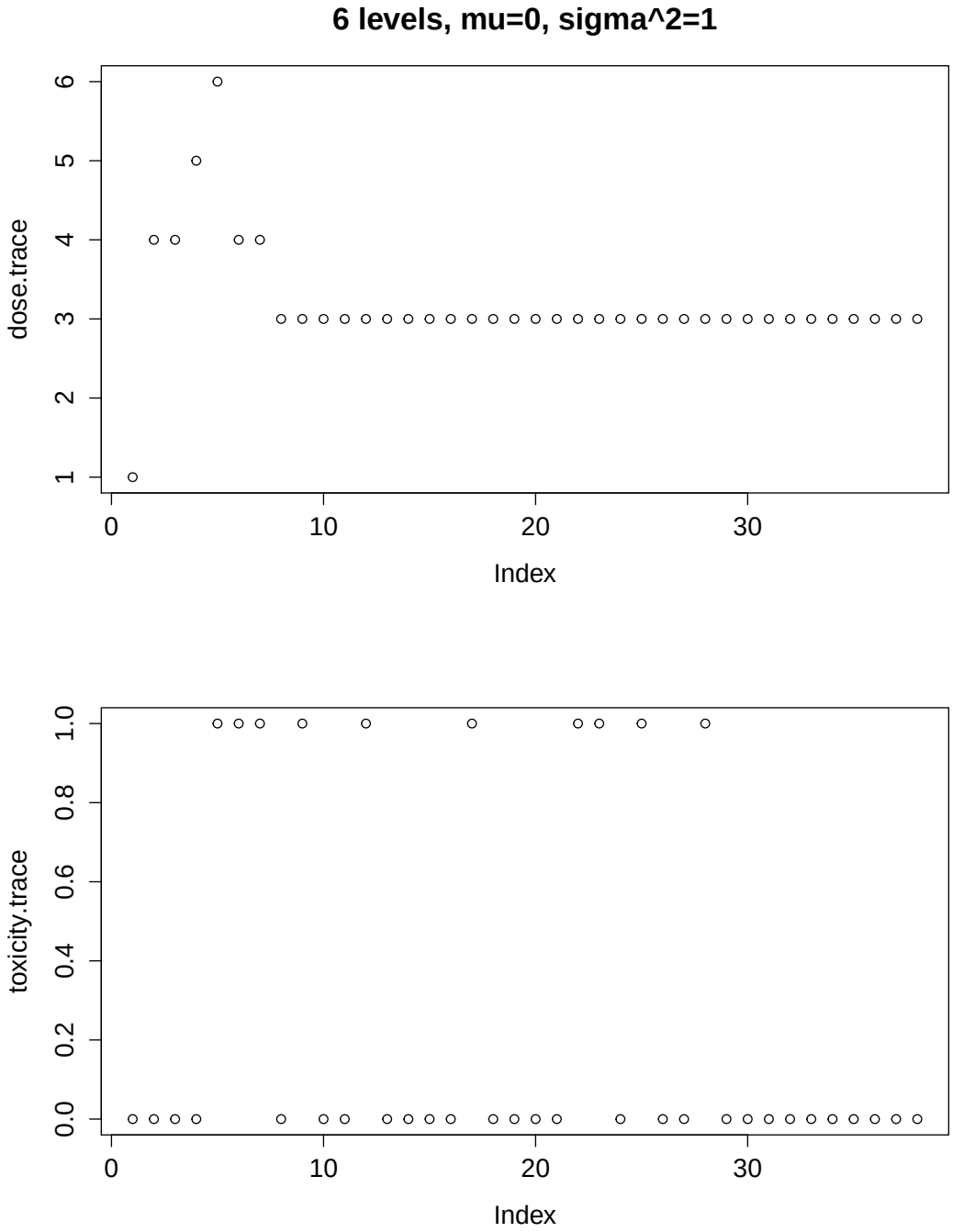


Figure B.19: Illustration of Dose Selection and Toxicities with 10 Dose Levels - Immediate Toxicity

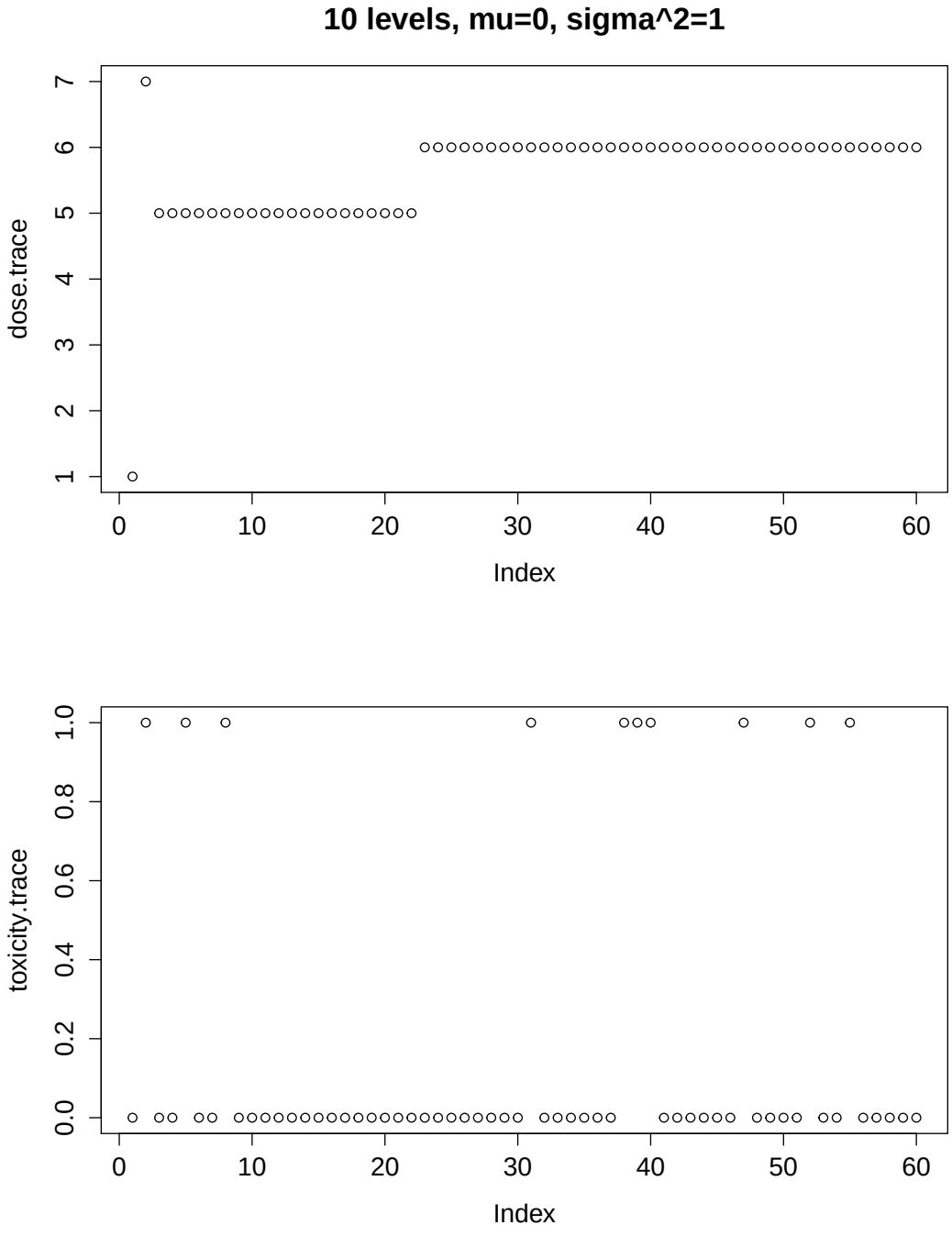


Figure B.20: Illustration of Dose Selection and Toxicities with 10 Dose Levels - Immediate Toxicity

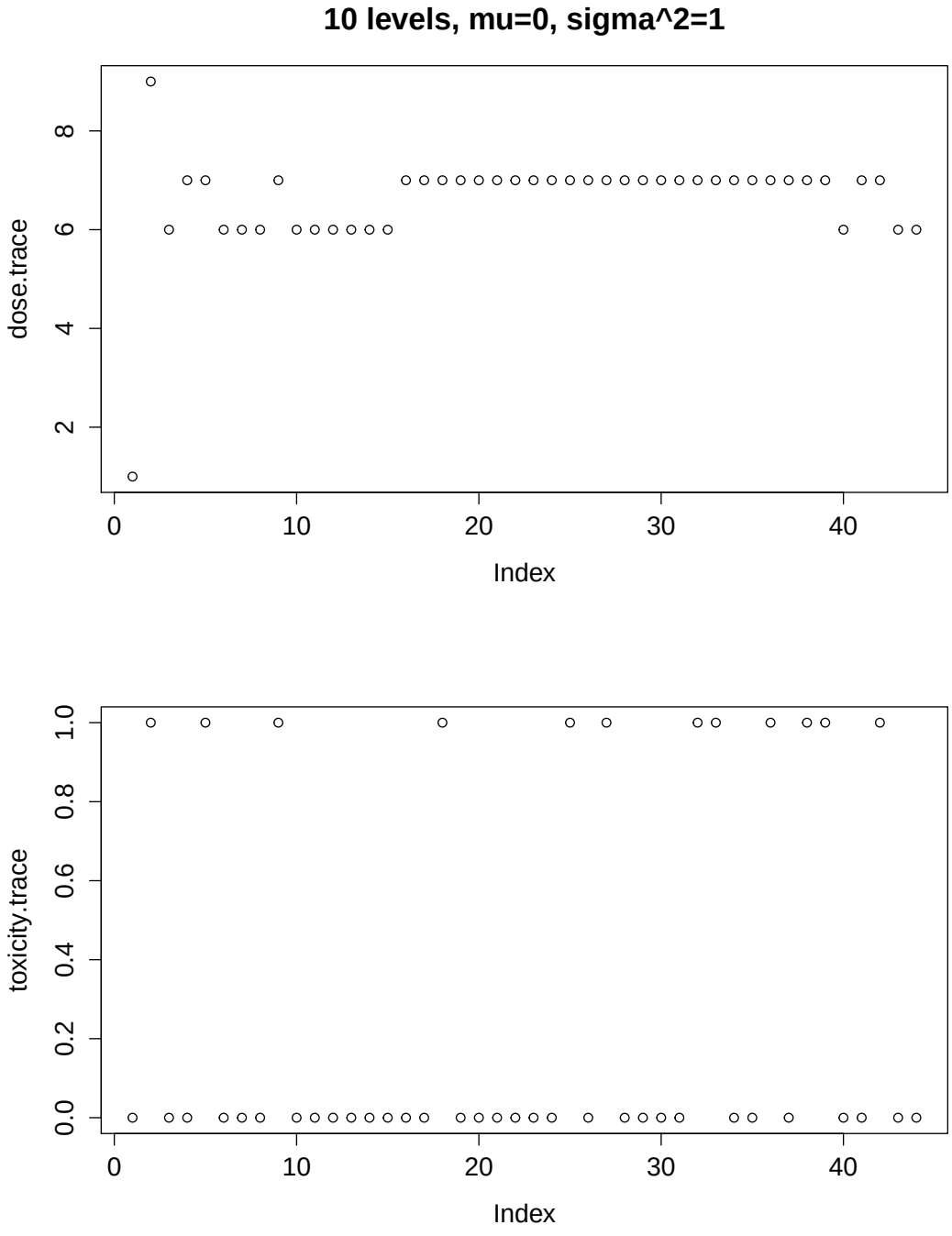


Figure B.21: Histogram of MTD for $N(0,1)$ and $Beta(2,5)$ prior distribution

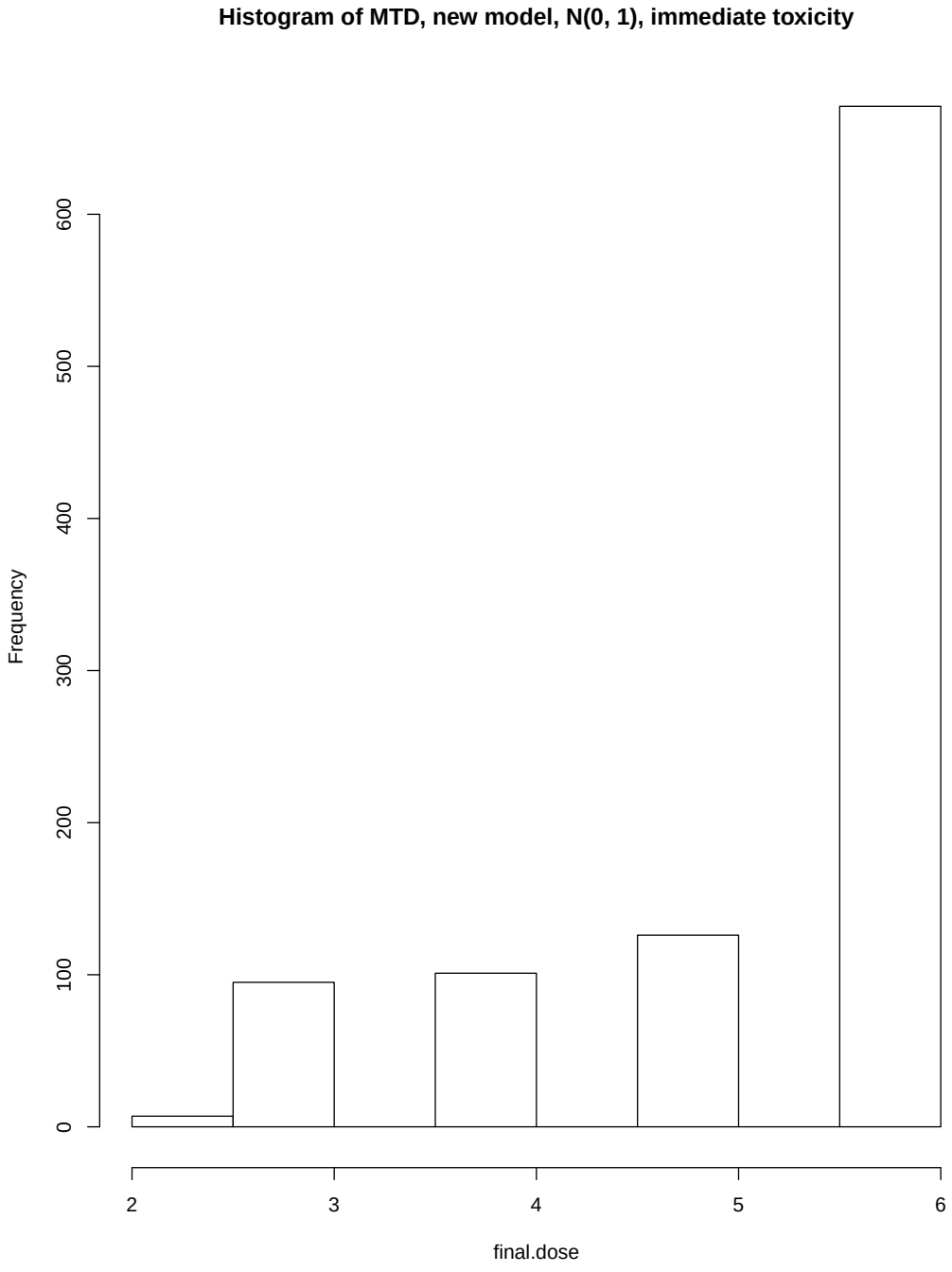


Figure B.22: Histogram of MTD for $N(0,1)$ and $Beta(2,10)$ prior distribution

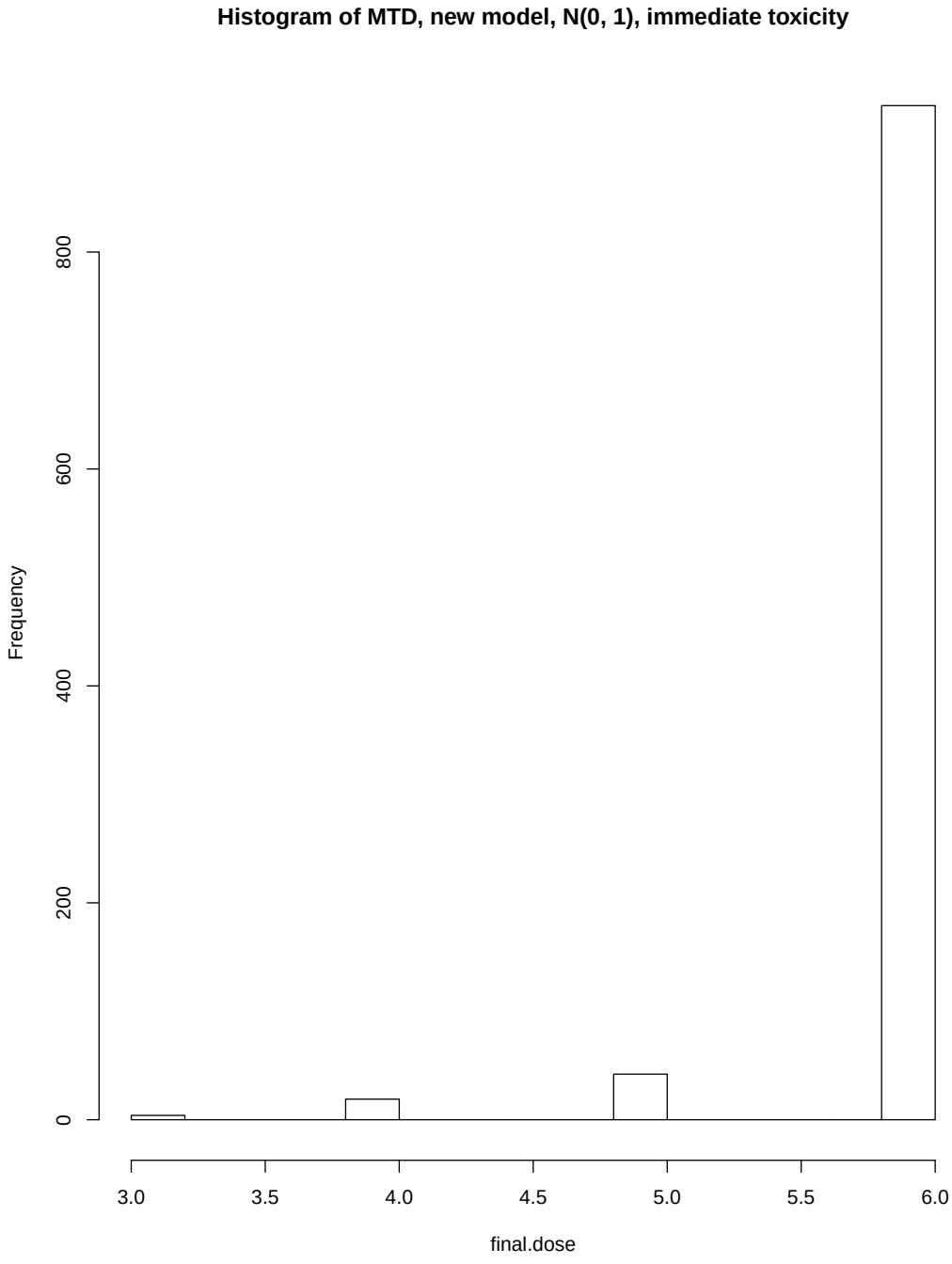


Figure B.23: Histogram of stopping times for $N(0,1)$ and $Beta(2,5)$ prior distribution

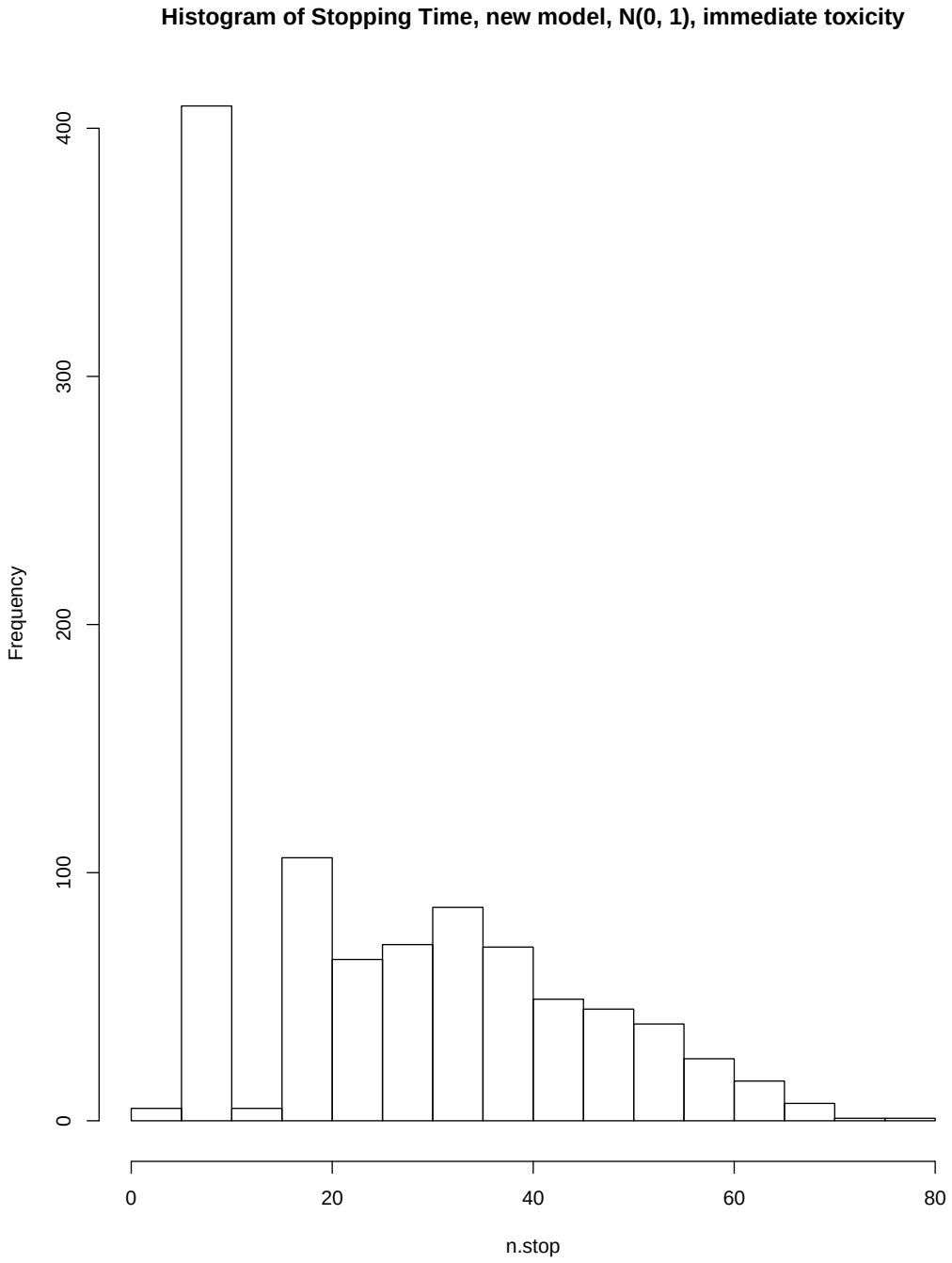


Figure B.24: Histogram of stopping times for $N(0,1)$ and $Beta(2,10)$ prior distribution

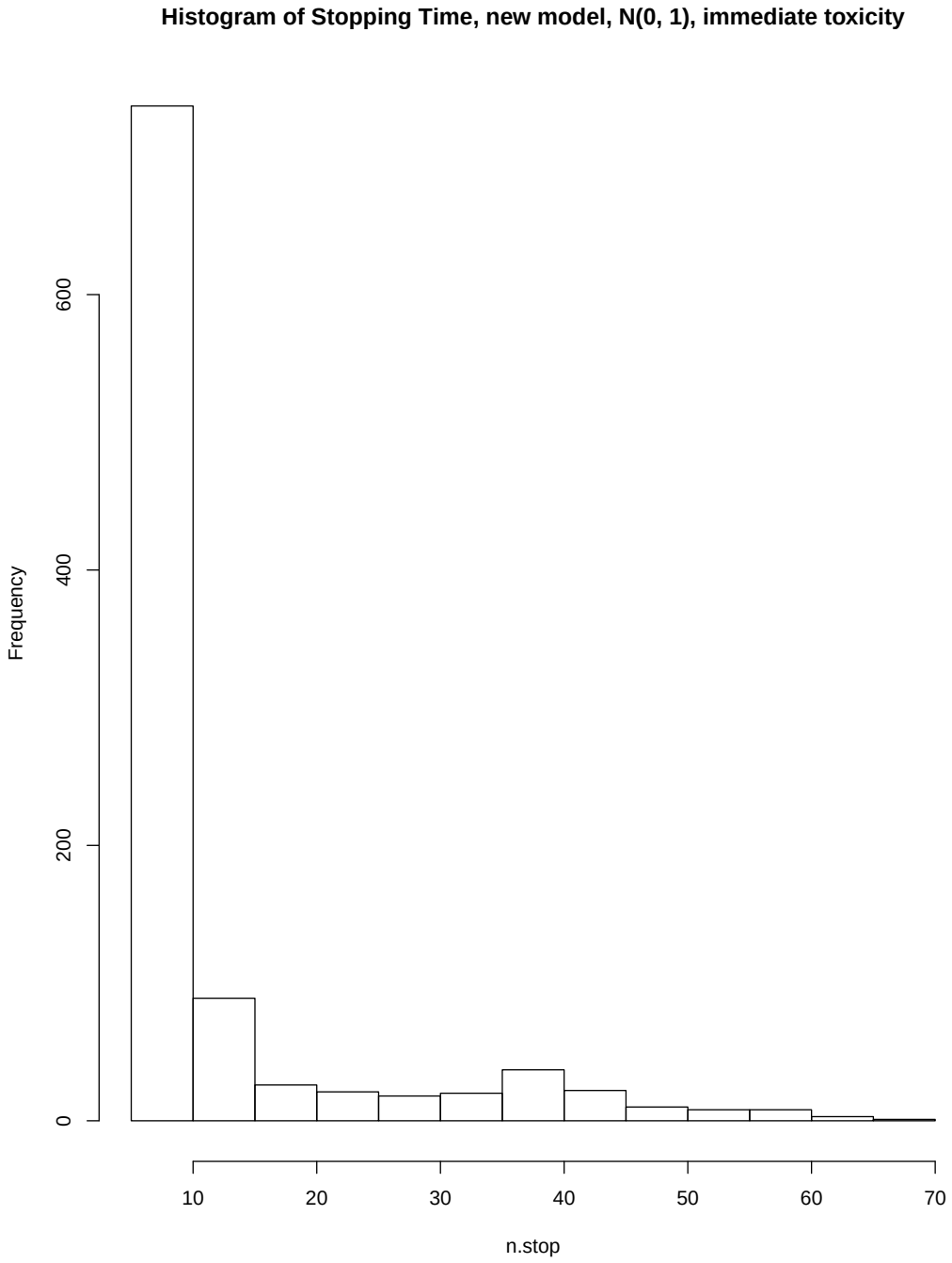


Figure B.25: Histogram of MTD for the uniform prior distribution

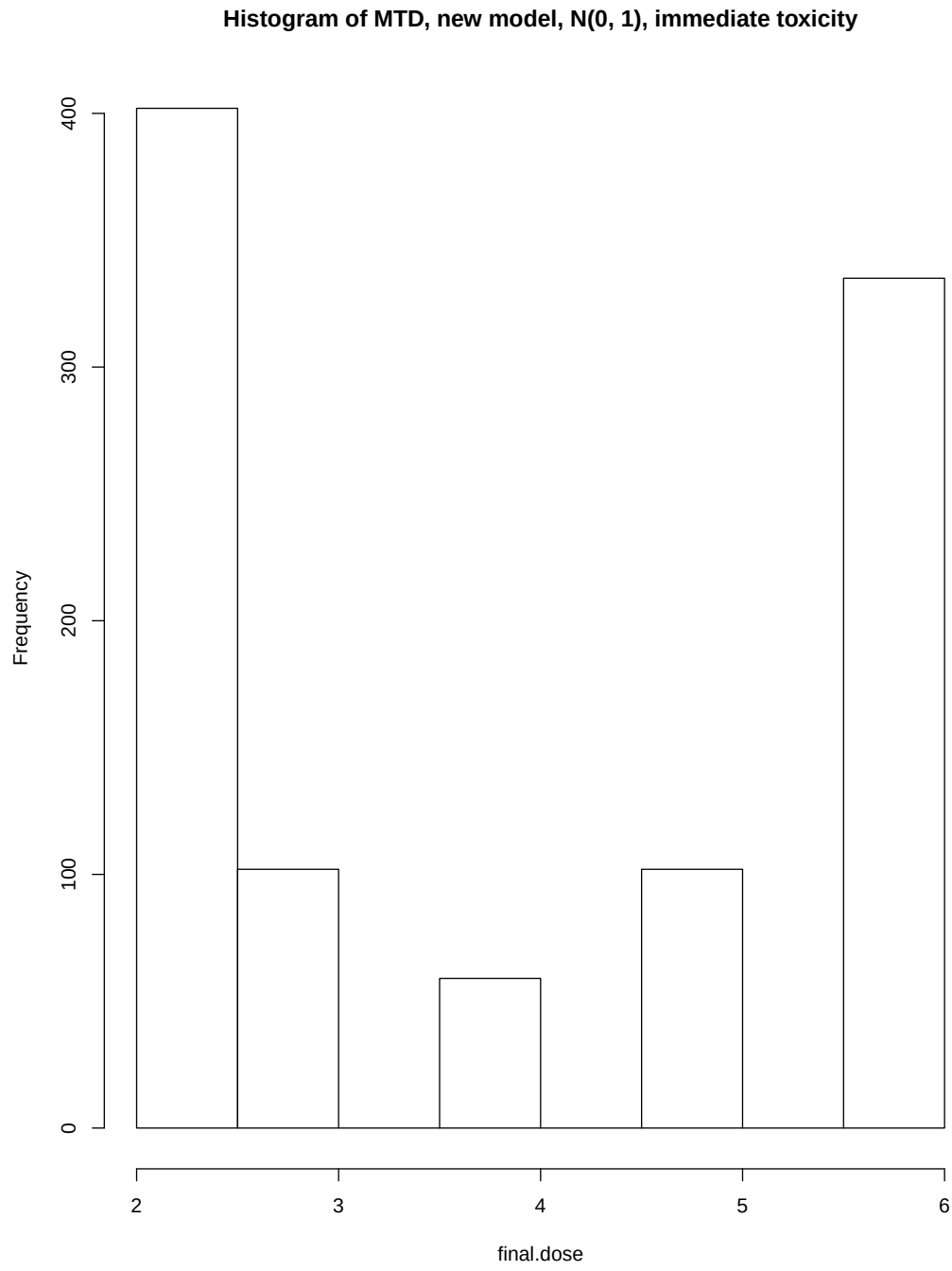
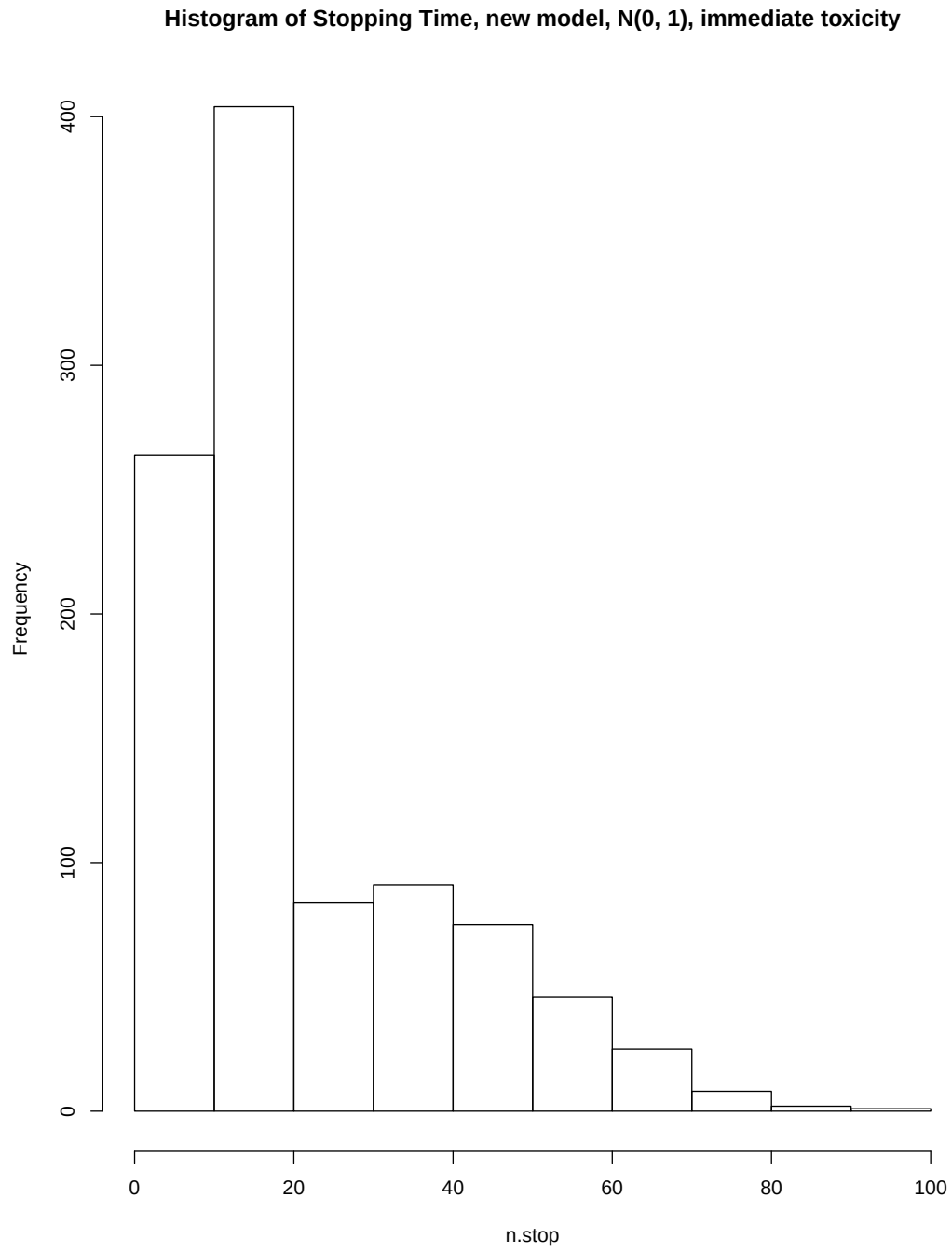


Figure B.26: Histogram of stopping times for the uniform prior distribution



Appendix C

Tables - immediate toxicity

Table C.1: Performance of different models

	Number at stop		Toxic Proportion		Dose probability	
Model	Mean	Variance	Mean	Variance	Mean	Variance
New, $\sigma^2 = 0.5$	19.329	150.081	0.143	0.017	0.143	0.002
New, $\sigma^2 = 1.0$	24.13	324.117	0.248	0.020	0.206	0.001
New, $\sigma^2 = 1.5$	32.086	304.211	0.272	0.010	0.253	0.0004
New, $\sigma^2 = 2.0$	29.386	269.985	0.323	0.016	0.280	0.0001
New, $\sigma^2 = 3.0$	2.000	0	0.338	0.111	0.337	0
Logistic	33.955	319.350	0.277	0.016	0.255	0.001
Hyperbolic	42.081	415.734	0.266	0.027	0.286	0.012

Table C.2: Toxicity probabilities of different models

	Dose					
Model	1	2	3	4	5	6
New, $\sigma^2 = 0.5$	0.000	0.002	0.055	0.221	0.393	0.528
New, $\sigma^2 = 1.0$	0.014	0.063	0.176	0.327	0.470	0.583
New, $\sigma^2 = 1.5$	0.094	0.178	0.290	0.410	0.520	0.611
New, $\sigma^2 = 2.0$	0.194	0.281	0.378	0.475	0.563	0.639
New, $\sigma^2 = 3.0$	0.337	0.404	0.471	0.536	0.596	0.650
Logistic	0.078	0.129	0.206	0.305	0.411	0.508
Hyperbolic	0.202	0.390	0.718	0.939	0.991	0.999

Appendix D

Graphs - late-onset toxicity

Table D.1: Boldface indicates toxicity, with number of days of late-onset observations in ( ), the italicized indicated no toxicity, and the underlined indicates convergence. The MTD is dose 6.

	Mean model toxicity probability at dose.					
Patient	1	2	3	4	5	6
1	*0.0145*	0.0638	0.1786	0.3315	0.4747	0.5880
2	0.0145	0.0633	0.1771	*0.3292*	0.4720	0.5853
3	0.0111	0.0416	0.1145	*0.2251*	0.3470	0.4590
4	0.0093	0.0308	0.0821	0.1664	*0.2696*	0.3744
5	0.0076	0.0215	0.0536	0.1102	1876	**0.2758** (1)
6	0.0089	0.0268	0.0688	0.1451	*0.2533*	0.3779
7	0.0083	0.0235	0.0583	0.1225	*0.2167*	0.3304
8	0.0078	0.0212	0.0511	0.1064	0.1897	**2938** (2)
9	0.0068	0.0165	0.0369	0.0743	0.1329	*0.2116*
10	0.0076	0.0198	0.0464	0.0958	0.1730	*0.2748*
11	0.0073	0.0183	0.0420	0.0856	0.1546	**0.2474** (0)
12	0.0081	0.0219	0.0526	0.1101	0.1995	*0.3158*
13	0.0078	0.0205	0.0482	0.1000	0.1815	*0.2896*
14	0.0075	0.0193	0.0447	0.0918	0.1666	**0.2673** (1)
15	0.0079	0.0209	0.0493	0.1026	0.1866	**0.2985** (5)
16	0.0075	0.0191	0.0438	0.0898	0.1632	*0.2633*
17	0.0072	0.0177	0.0396	0.0801	0.1451	*0.2355*
18	0.0069	0.0166	0.0363	0.0725	0.1308	*0.2130*
19	0.0067	0.0156	0.0337	0.0663	0.1192	*0.1944*
20	0.0070	0.0169	0.0371	0.0743	0.1343	*0.2189*
21	0.0069	0.0164	0.0359	0.0715	0.1290	*0.2105*
22	0.0068	0.0161	0.0348	0.0689	0.1241	*0.2026*
23	0.0067	0.0157	0.0338	0.0666	0.1196	*0.1954*
24	0.0067	0.0154	0.0328	0.0644	0.1154	*0.1886*
25	0.0066	0.0151	0.0319	0.0624	0.1115	**0.1823** (1)
26	0.0068	0.0160	0.0347	0.0687	0.1237	*0.2022*
27	0.0068	0.0157	0.0338	0.0667	0.1199	**0.1961** (1)
28	0.0070	0.0166	0.0363	0.0723	0.1305	*0.2135*
29	0.0069	0.0163	0.0354	0.0703	0.1268	*0.2076*
30	0.0068	0.0160	0.0346	0.0685	0.1233	*0.2019*
31	0.0068	0.0158	0.0339	0.0668	0.1201	**0.1966** (1)
32	0.0069	0.0165	0.0360	0.0716	0.1294	*0.2118*
33	0.0069	0.0163	0.0353	0.0699	0.1262	*0.2066*
34	0.0068	0.0160	0.0346	0.0684	0.1231	*0.2017*

Figure D.1: Illustration of Dose Trace with 6 Dose Levels - Late-onset Toxicity

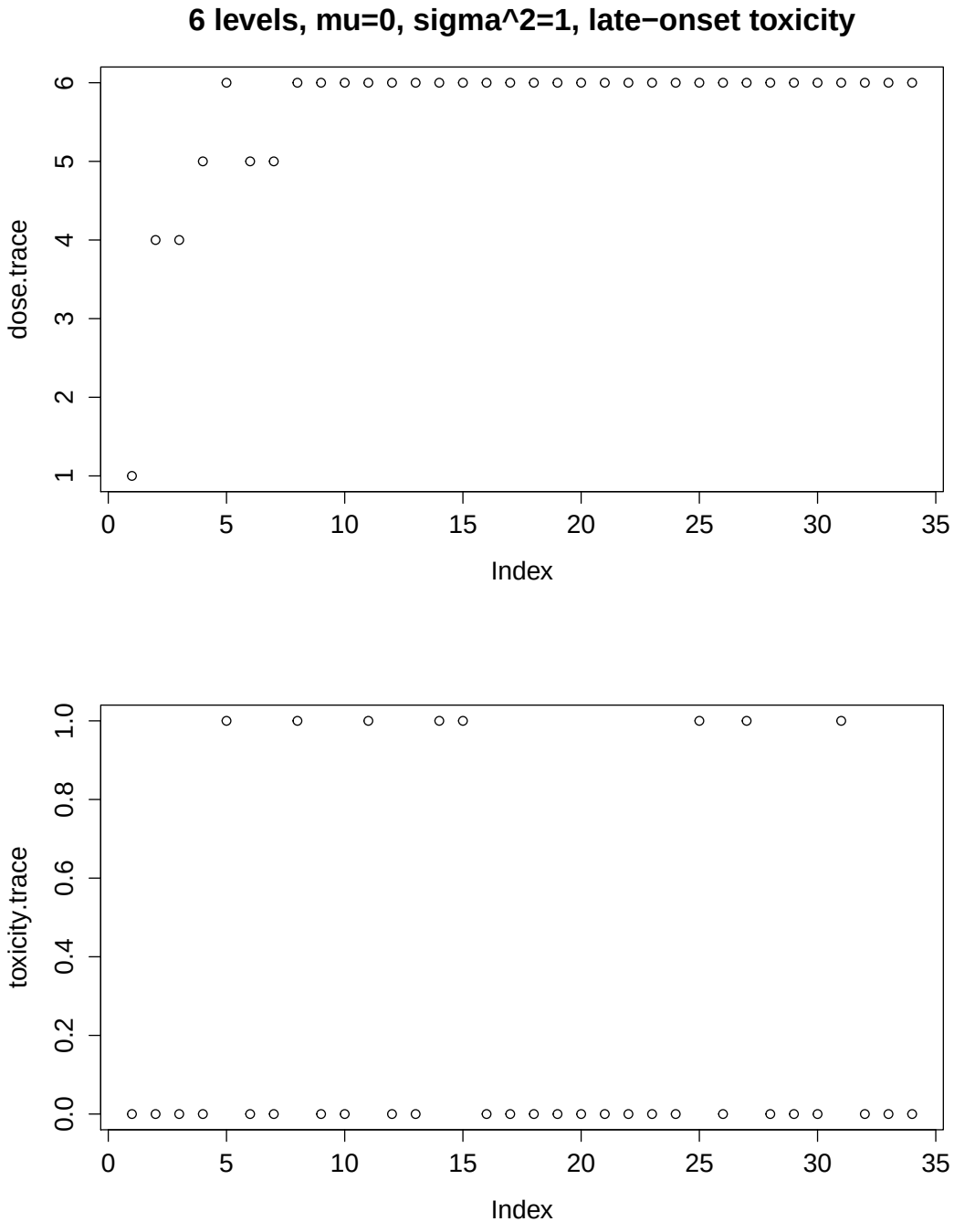
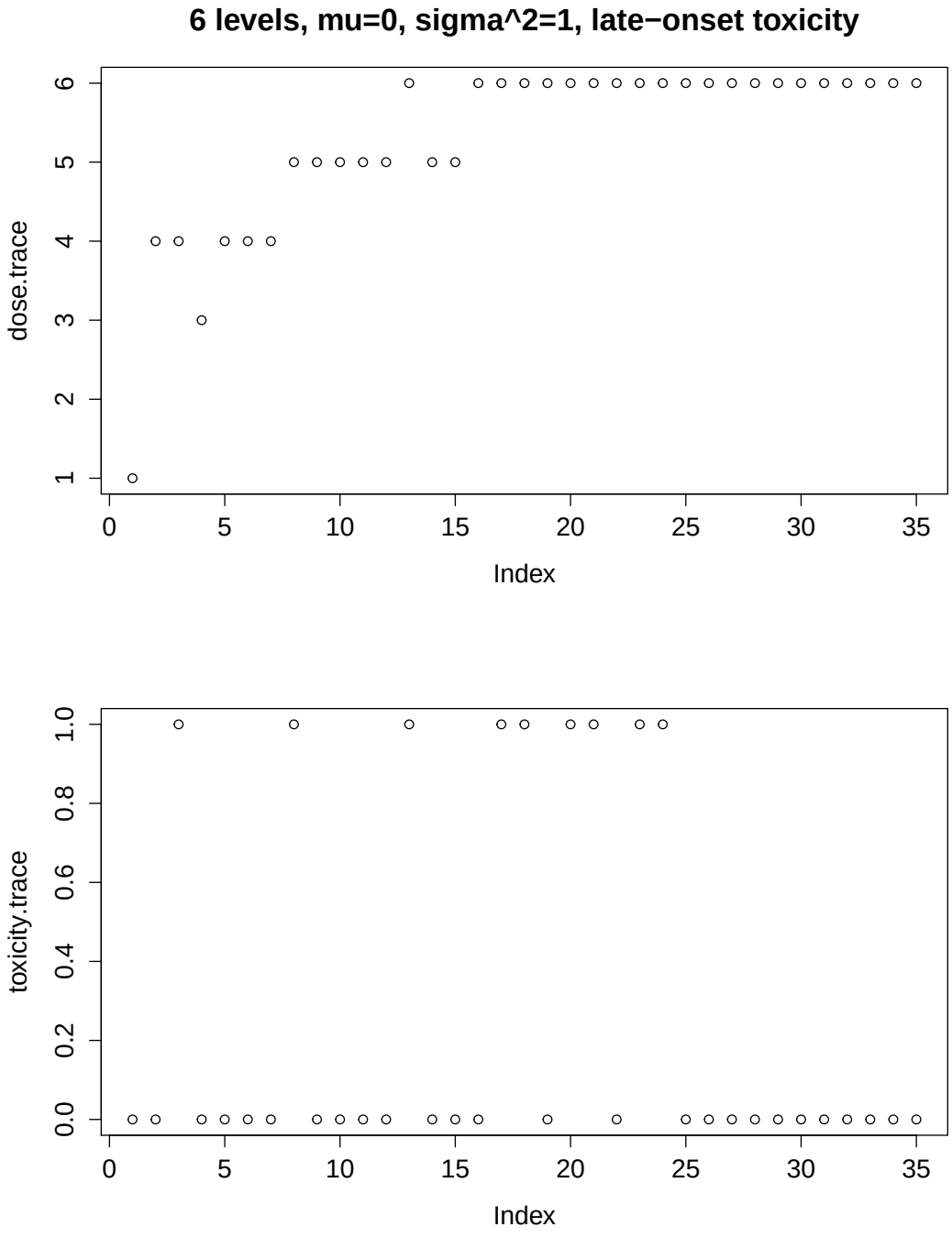


Figure D.2: Illustration of Dose Trace with 6 Dose Levels - Late-onset Toxicity



Appendix E

Tables - late-onset toxicity

Table E.1: Performance of different models

	Number at stop		Toxic Proportion		Dose probability	
Model	Mean	Variance	Mean	Variance	Mean	Variance
New, $\sigma^2 = 0.5$	17.672	92.573	0.115	0.010	0.126	0.001
New, $\sigma^2 = 1.0$	24.85	152.86	0.184	0.013	0.183	0.002
New, $\sigma^2 = 2.0$	21.286	155.55	0.299	0.017	0.278	0.0001
Logistic	23.064	133.658	0.231	0.016	0.218	0.001
Hyperbolic	19.078	44.472	0.237	0.017	0.235	0.0003

Table E.2: Toxicity probabilities of different models

	Dose					
Model	1	2	3	4	5	6
New, $\sigma^2 = 0.5$	0.000	0.0002	0.008	0.052	0.129	0.230
New, $\sigma^2 = 1.0$	0.007	0.020	0.048	0.092	0.148	0.212
New, $\sigma^2 = 2.0$	0.156	0.193	0.234	0.278	0.321	0.364
Logistic	0.060	0.076	0.098	0.126	0.159	0.196
Hyperbolic	0.065	0.224	0.608	0.912	0.987	0.998

Figure E.1: Histogram of final selection of MTD, new model, late-onset toxicity

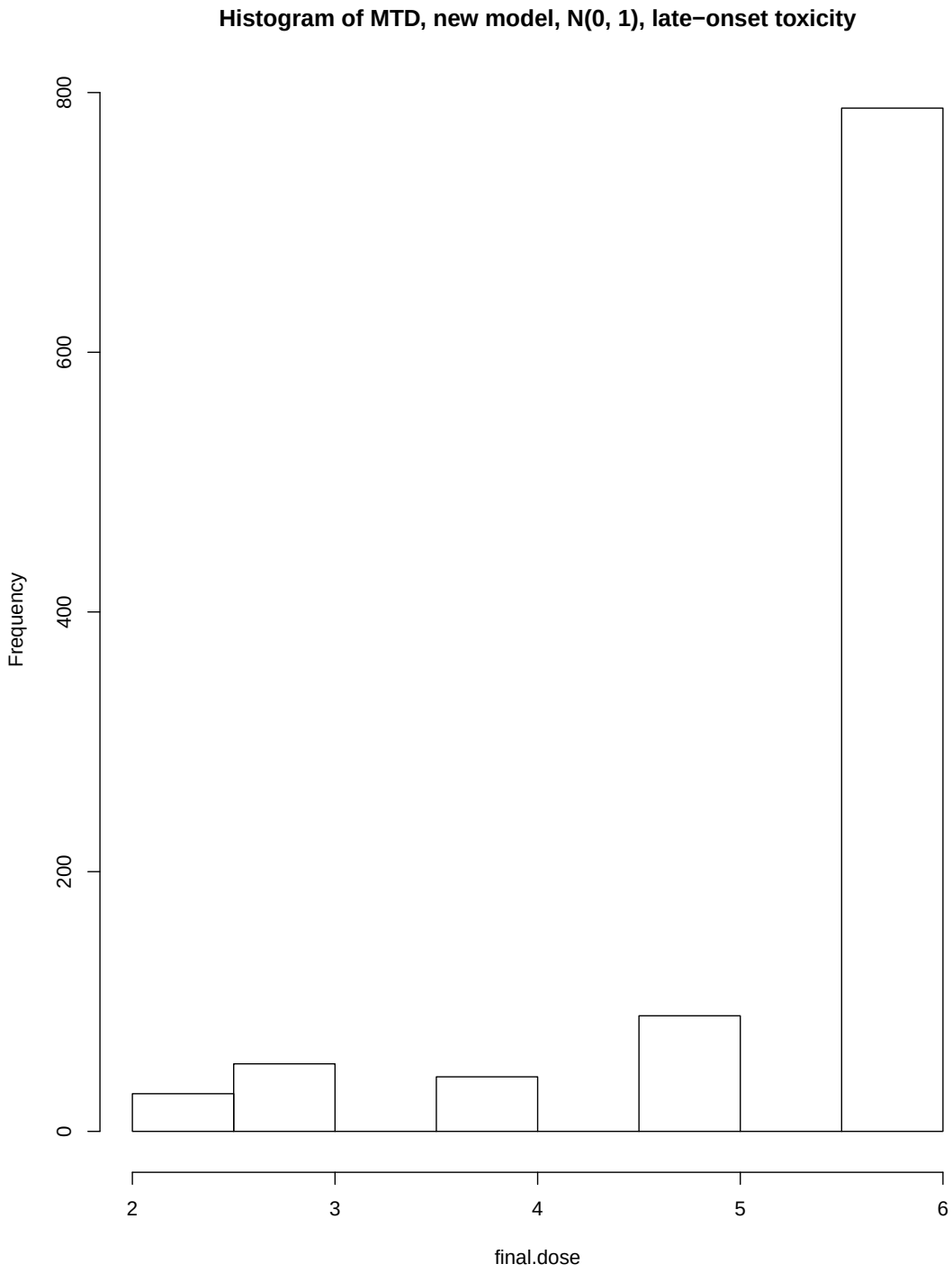


Figure E.2: Histogram of stopping times, new model, late-onset toxicity

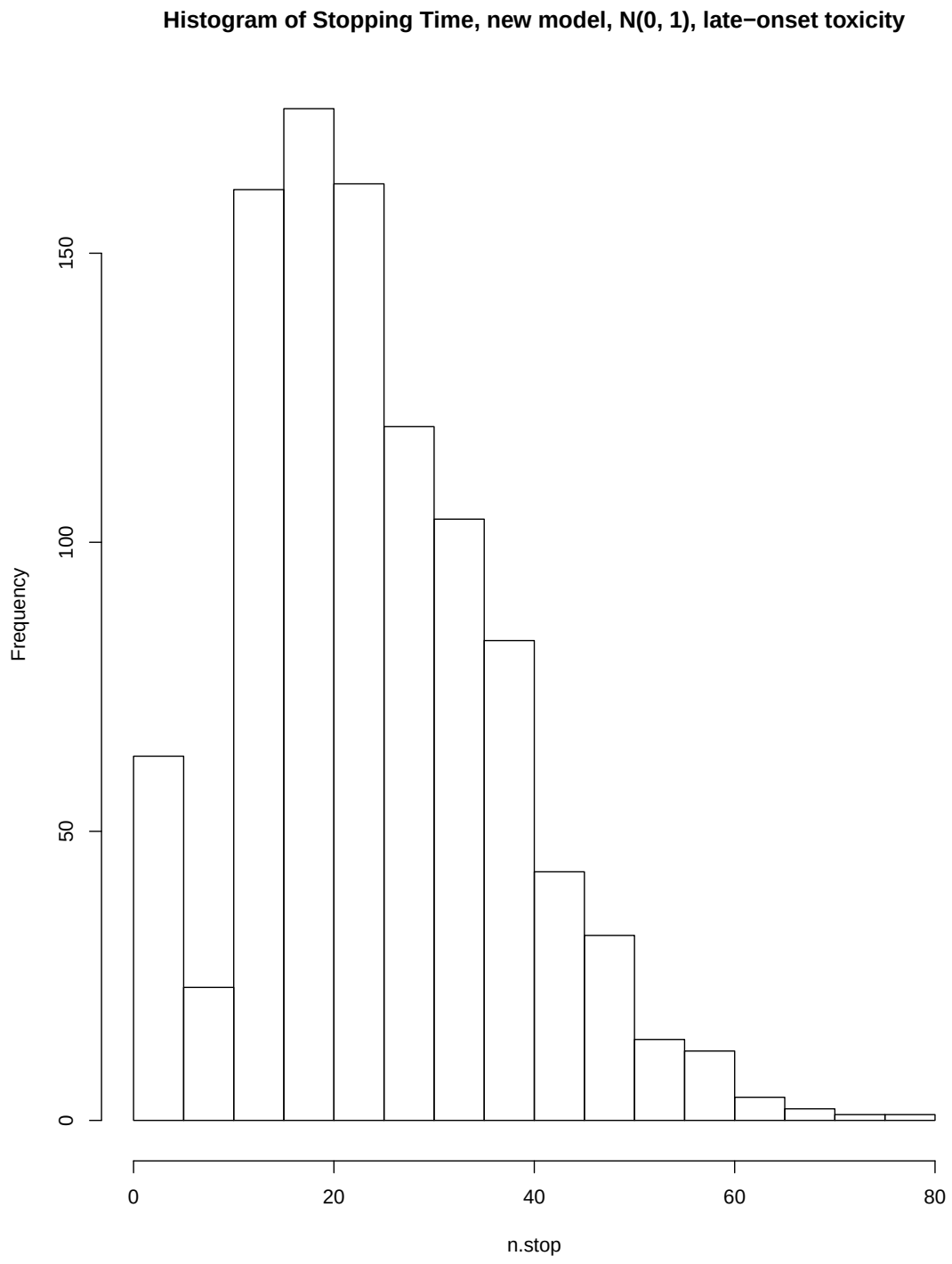


Figure E.3: Histogram of final selection of MTD, new model, late-onset toxicity

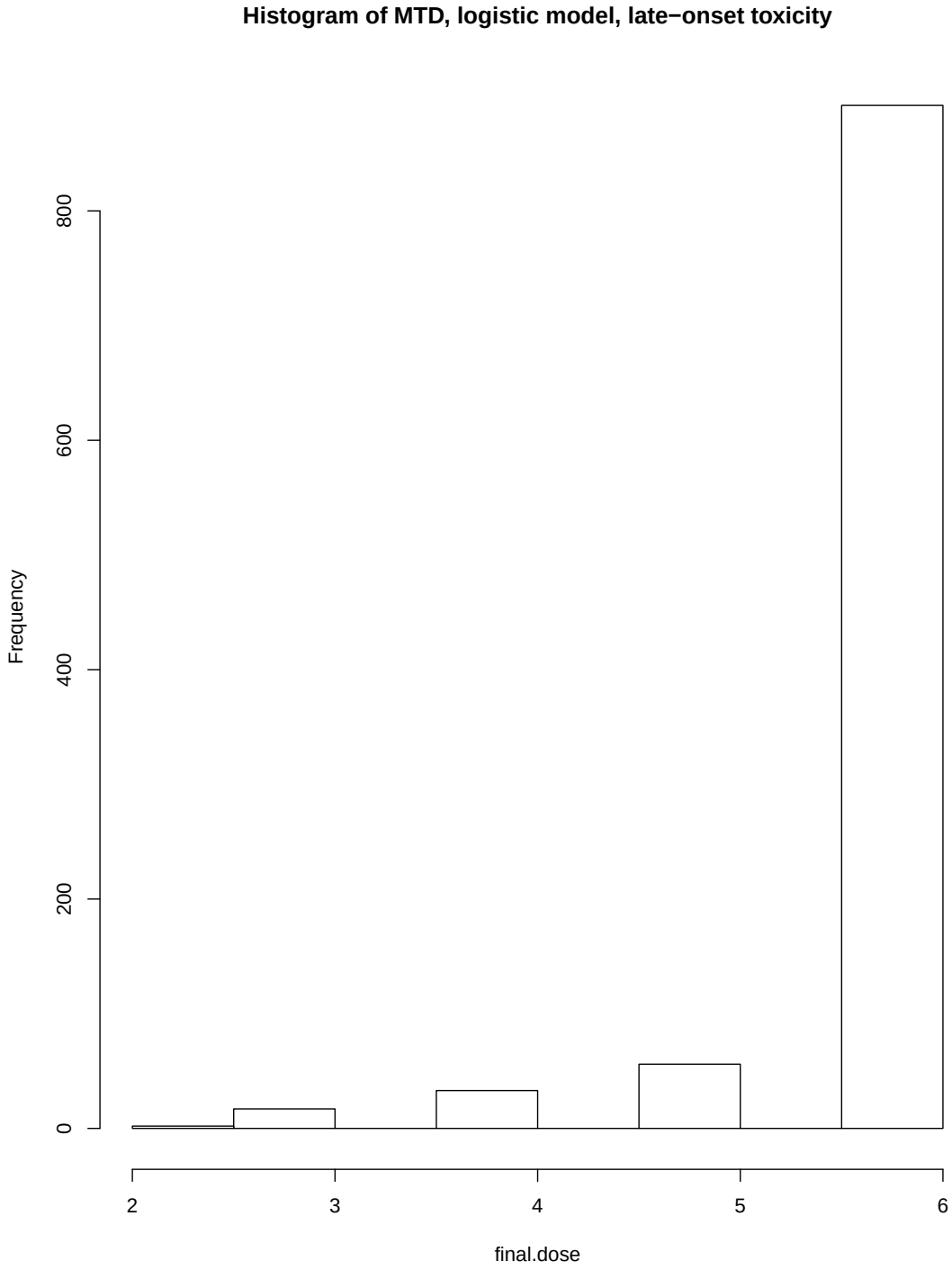


Figure E.4: Histogram of stopping times, new model, late-onset toxicity

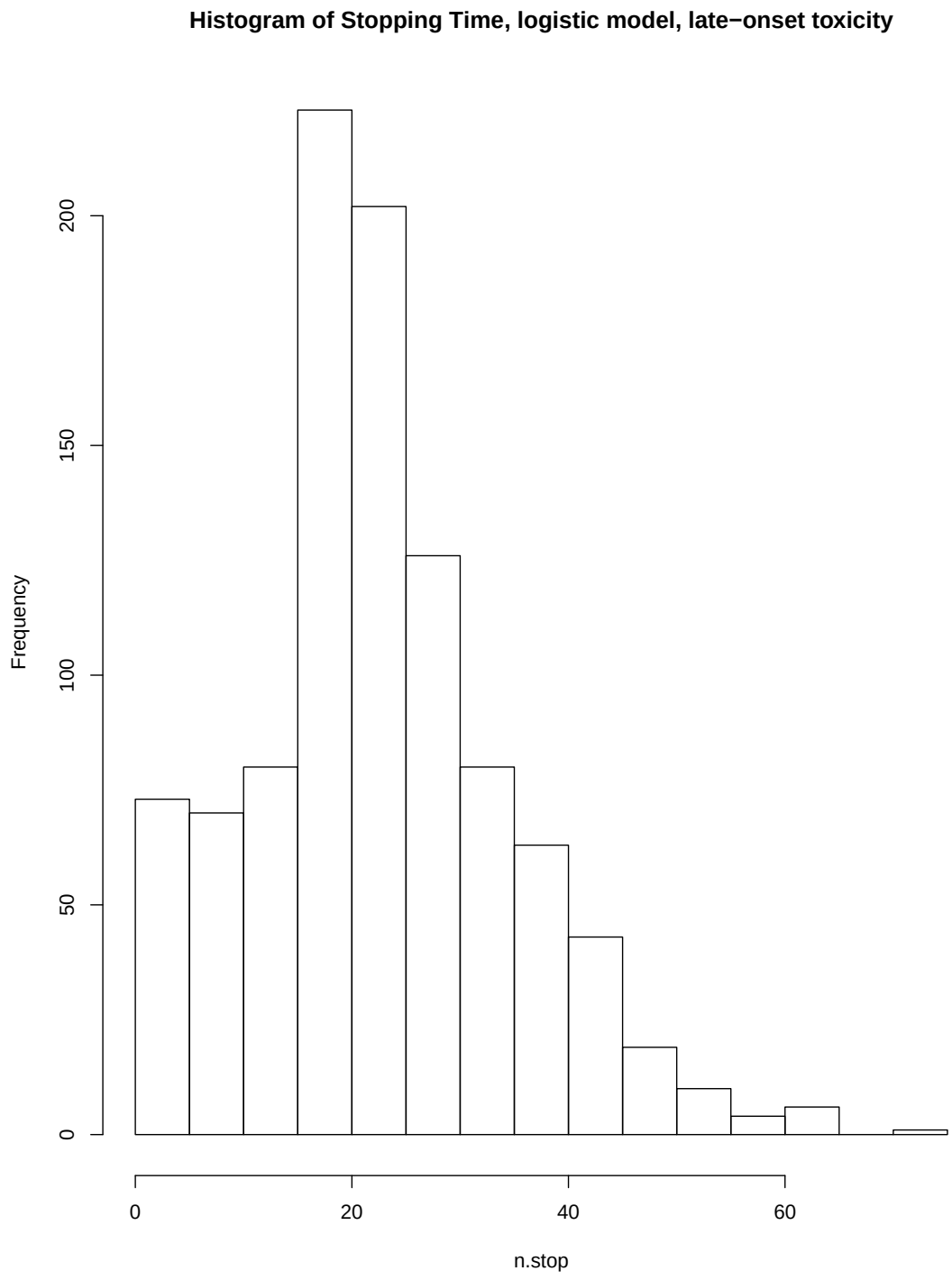


Figure E.5: Histogram of final selection of MTD, new model, late-onset toxicity

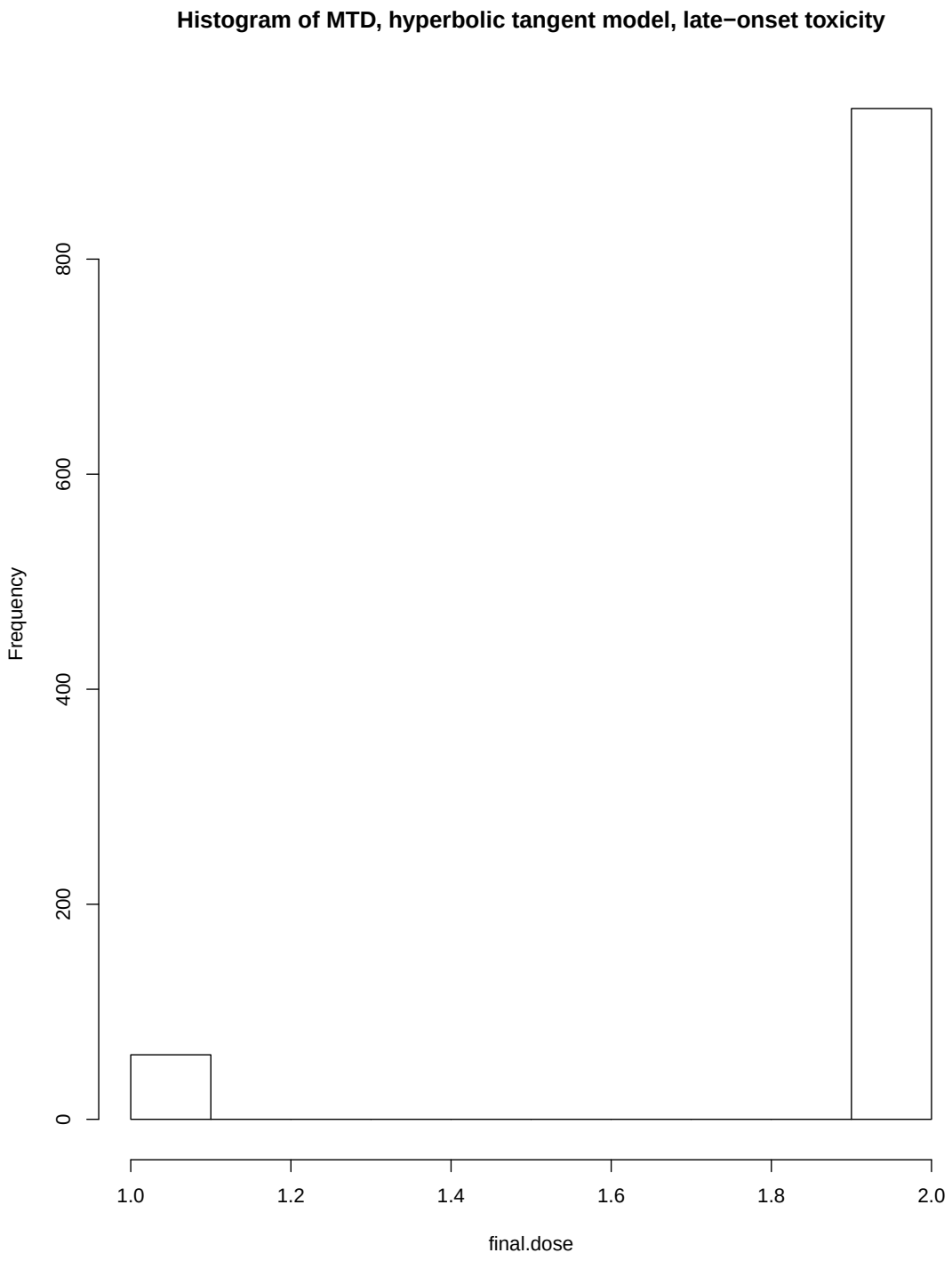


Figure E.6: Histogram of stopping times, new model, late-onset toxicity