
A New Approach to Dose Estimation in Drug Development Based on Maximization of Likelihood of Grouped Data

Nicholas A. Nechval^{1, *}, Gundars Berzins², Vadims Danovics³

¹Department of Mathematics, Baltic International Academy, Riga, Latvia

²Department of Management, University of Latvia, Riga, Latvia

³Department of Marketing, University of Latvia, Riga, Latvia

Email address:

nechval@junik.lv (N. A. Nechval), gundars.berzins@lu.lv (G. Berzins), vadims.danovics@lu.lv (V. Danovics)

To cite this article:

Nicholas A. Nechval, Gundars Berzins, Vadims Danovics. A New Approach to Dose Estimation in Drug Development Based on Maximization of Likelihood of Grouped Data. *American Journal of Theoretical and Applied Statistics*. Special Issue: Novel Ideas for Efficient Optimization of Statistical Decisions and Predictive Inferences under Parametric Uncertainty of Underlying Models with Applications. Vol. 5, No. 2-1, 2016, pp. 12-20. doi: 10.11648/j.ajtas.s.2016050201.13

Abstract: Identifying the ‘right’ dose is one of the most critical and difficult steps in the clinical development process of any medicinal drug. Its importance cannot be understated: selecting too high a dose can result in unacceptable toxicity and associated safety problems, while choosing too low a dose leads to smaller chances of showing sufficient efficacy in confirmatory trials, thus reducing the chance of approval for the drug. The optimal dose is the dose that gives the desired effect with minimum side effects. The dose of a drug is of course ‘optimal’ only for a given subject, but not necessarily for any other. In view of this the objective of a dose-finding trials is not to determine a single fixed dose for use in the early phases of clinical trials or in medical practice, but to determine an interval of doses within which there is a stated degree of confidence that the defined, acceptable therapeutic response and the frequency of adverse reactions will lie above and below, respectively, certain acceptable predetermined levels. If the subject samples used in the dose finding studies adequately represent the subject population for which the drug is intended, the interval of doses so defined can be applied to the subject population as a whole. In this paper, we propose the technique based on maximization of likelihood function in order to estimate the maximal tolerated dose (MTD) and minimal effective dose (MED) on the basis of l samples of subjects, which are grouped in a simplest way. The necessary and sufficient conditions for the existence and uniqueness of the maximum likelihood estimates are derived. The proposed approach to dose estimation in drug development is simple and suitable for medical practice. The numerical examples are given.

Keywords: Drug Development, Dose Estimation, Grouped Data, Likelihood Function, Maximization

1. Introduction

The proper understanding and characterization of the dose-response relationship for a new compound is a fundamental step in the clinical development process of any medicinal drug. The determination of the doses of a new pharmaceutical preparation to be used in clinical practice is a very important issue in the early phases of clinical trials: this is recognized not only in the literature [1] but is also mentioned in the F.D.A. guidelines for the clinical evaluation of drugs [2].

A common problem in toxicological and drug development studies is to assess the biological activity of a chemical compound. For this purpose, a dose-response experiment is conducted in which several doses of the compound are administered to separate groups of experimental units. There are two primary goals in these studies. In a toxicological study

the goal is to estimate a safe dose that will not cause some undesirable effect (e.g., toxicity, carcinogenicity), whereas in a drug development study the goal is to estimate the lowest dose that will cause some desirable effect. Indeed, in Phase I clinical trials, researchers test a new drug or treatment in a small group of subjects for the first time in order to evaluate its safety, identify side-effects and determine a therapeutically useful interval of doses. The upper end of the interval is the maximal tolerated dose (MTD) and the lower end of the interval is the minimal effective dose (MED). Here we will face the estimation of the minimal effective dose; generalization to maximum tolerated dose will be shown straightforward.

A well-established approach to search for MED and MTD is based on the construction of a dose-toxicity curve. Robbins and Monro [3] and later Wu [4] try to estimate, through

stochastic approximation, the MTD as the quantile of this curve. Eichhorn and Zacks [5, 6] studied the sequential search problem through linear regression dose-toxicity models. Inference within the dose-response framework that accounts for the model uncertainty was discussed by Pinheiro, Bretz, and Branson [7]; Bornkamp, Pinheiro, and Bretz [8]; and Whitney and Ryan [9] who used parametric nonlinear models to characterize the dose-response relationship and proposed to use model averaging techniques to account for model uncertainty. Within this approach several parametric nonlinear models are fitted to the data and information from all models is combined, using information criteria, for both estimation and inference. Although some of these methods perform quite well in small sample dose-response study simulations, their main drawback is just requiring an explicit dose-toxicity curve, often nonparametric, that could be artificial and complicated. Gasparini and Eisele [10] proposed a curve-free method: modelling the probabilities of toxicity directly as the unknown parameter of interest, a product of beta prior distributions is considered.

In this paper, the parameters of interest are the probability distribution functions of a suitable dose in drug development (from the point of view of toxicity and from the point of view of efficacy, respectively), which are determined via a new approach based on maximization of likelihood of grouped data.

2. Minimal Effective Dose Estimation

2.1. Likelihood Function of Grouped Data

Let us assume that the random variable X , which represents an effective dose level of a drug (from the point of view of efficacy) for randomly chosen subject, has a continuous cumulative distribution function $F_e(x | \theta_e)$ (probability density function $f_e(x | \theta_e)$) with unknown parametric vector θ_e . We consider l ($l \geq 1$) the number of components of the unknown parametric vector θ_e) random samples of subjects of sizes $N_j, j=1(1)l$. The N_j subjects of the j th random sample are assigned to the dose d_j of a drug. Let n_j be the number of subjects in the j th sample, for which the effective dose of a drug is less than d_j . It is assumed, without loss of generality, that $0 < d_1 < d_2 < \dots < d_l$.

The problem is to estimate the unknown parametric vector θ_e . For this purpose, it can be used the likelihood function of the grouped data

$$L_e(\theta_e) = \prod_{j=1}^l \binom{N_j}{n_j} [\Pr(X \leq d_j)]^{n_j} [\Pr(X > d_j)]^{N_j - n_j} \\ = \prod_{j=1}^l \binom{N_j}{n_j} F_e^{n_j}(d_j | \theta_e) [1 - F_e(d_j | \theta_e)]^{N_j - n_j}. \quad (1)$$

Consider a situation described by a location-scale family of probability distribution functions, indexed by a parametric vector $\theta_e = (\mu_e, \sigma_e)$,

$$F_e(x | \theta_e) = \Pr(X \leq x) = F_e\left(\frac{x - \mu_e}{\sigma_e}\right) = F_e(z), \quad -\infty < x < \infty, \quad (2)$$

where $-\infty < \mu_e < \infty$ is a location parameter and $\sigma_e > 0$ is a scale parameter, the distribution of $Z = (X - \mu_e) / \sigma_e$ does not depend on any unknown parameters.

Assumption 1. $F_e(z)$ is strictly increasing continuous function for all $z \in (-\infty, \infty)$, i.e., $F_e(z_2) > F_e(z_1)$ for all $z_2 > z_1$.

For computational convenience, the maximum likelihood estimate is obtained by maximizing the log-likelihood function, $\ln L_e(\theta_e)$. This is because the two functions, $L_e(\theta_e)$ and $\ln L_e(\theta_e)$, are monotonically related to each other so the same maximum likelihood (ML) estimate is obtained by maximizing either one. Thus,

$$\ln L_e(\theta_e) = \sum_{j=1}^l \ln \binom{N_j}{n_j} + \sum_{j=1}^l n_j \ln F_e(d_j | \theta_e) \\ + \sum_{j=1}^l (N_j - n_j) \ln [1 - F_e(d_j | \theta_e)] \\ = \sum_{j=1}^l \ln \binom{N_j}{n_j} + \sum_{j=1}^l n_j \ln F_e(z_j) \\ + \sum_{j=1}^l (N_j - n_j) \ln [1 - F_e(z_j)] \quad (3)$$

where

$$z_j = \frac{d_j - \mu_e}{\sigma_e}. \quad (4)$$

2.2. Log-Likelihood Equations

To obtain the log-likelihood equations for the unknown parameters μ_e and σ_e , the log-likelihood function in Equation (3) will be differentiated partially with respect to μ_e and σ_e , respectively, and set equal to zero to provide the maximum likelihood estimates (MLEs), $\hat{\mu}_e$ and $\hat{\sigma}_e$, of the unknown parameters, μ_e and σ_e , respectively, as follows:

$$\frac{\partial \ln L_e(\theta_e)}{\partial \mu_e} = \frac{1}{\sigma_e} \sum_{j=1}^l [(N_j - n_j) h_{e1}(z_j) - n_j h_{e2}(z_j)] = 0, \quad (5)$$

$$\frac{\partial \ln L_e(\theta_e)}{\partial \sigma_e} = \frac{1}{\sigma_e} \sum_{j=1}^l z_j [(N_j - n_j) h_{e1}(z_j) - n_j h_{e2}(z_j)] = 0, \quad (6)$$

where

$$h_{e1}(z_j) = \frac{F'_e(z_j)}{1 - F_e(z_j)}, \quad h_{e2}(z_j) = \frac{F'_e(z_j)}{F_e(z_j)}. \quad (7)$$

Equations (5) and (6), which represent a necessary condition for the existence of the MLEs, can be reduced to:

$$\sum_{j=1}^l (N_j - n_j) h_{e1}(z_j) - \sum_{j=1}^l n_j h_{e2}(z_j) = 0, \quad (8)$$

$$\sum_{j=1}^l d_j (N_j - n_j) h_{e1}(z_j) - \sum_{j=1}^l d_j n_j h_{e2}(z_j) = 0. \quad (9)$$

It is usually not possible to obtain an analytic form solution of Equations (8) and (9). Hence, a numerical iterative technique can be used to obtain the MLEs, $\hat{\mu}_e$ and $\hat{\sigma}_e$, of the parameters, μ_e and σ_e , respectively.

2.3. Existence and Uniqueness of the ML Estimates

The maximum likelihood estimates need not exist nor be unique. In this section, we show that the MLEs, $\hat{\mu}_e$ and $\hat{\sigma}_e$, which maximize (1), exist in the above case and are unique.

Assumption 2. $F_e'(z)$ and $F_e''(z)$ exist and are continuous functions for all $z \in (-\infty, \infty)$.

Assumption 3. $h_{e1}(z)$ is strictly increasing function for all $z \in (-\infty, \infty)$; $h_{e2}(z)$ is strictly decreasing function for all $z \in (-\infty, \infty)$.

It follows from (8) and (9) that the MLEs, $\hat{\mu}_e$ and $\hat{\sigma}_e$, exist and are unique if and only if

$$\sum_{j=1}^l (N_j - n_j) \sum_{j=1}^l d_j n_j > \sum_{j=1}^l d_j (N_j - n_j) \sum_{j=1}^l n_j \quad (10)$$

or

$$\frac{\sum_{j=1}^l d_j n_j}{\sum_{j=1}^l d_j N_j} > \frac{\sum_{j=1}^l n_j}{\sum_{j=1}^l N_j}. \quad (11)$$

A sufficient condition for $\hat{\theta}_e = (\hat{\mu}_e, \hat{\sigma}_e)$ to be a maximum of (3) is that the Hessian matrix

$$\mathbf{H}(\theta_e) = \begin{bmatrix} \frac{\partial^2 \ln L_e(\theta_e)}{\partial \mu_e^2} & \frac{\partial^2 \ln L_e(\theta_e)}{\partial \mu_e \partial \sigma_e} \\ \frac{\partial^2 \ln L_e(\theta_e)}{\partial \sigma_e \partial \mu_e} & \frac{\partial^2 \ln L_e(\theta_e)}{\partial \sigma_e^2} \end{bmatrix} \quad (12)$$

evaluated at $\theta_e = \hat{\theta}_e$ satisfies the following condition [11]:

$$\mathbf{u}' \mathbf{H}(\theta_e) \mathbf{u} < 0, \quad (13)$$

where

$$\mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \quad (14)$$

It follows from (5) and (6) that

$$\frac{\partial^2 \ln L_e(\theta_e)}{\partial \mu_e^2} = -\frac{1}{\sigma_e^2} \sum_{j=1}^l Q_j, \quad (15)$$

where

$$Q_j = (N_j - n_j) h'_{e1}(z_j) - n_j h'_{e2}(z_j) > 0, \quad (16)$$

$$\frac{\partial^2 \ln L_e(\theta_e)}{\partial \mu_e \partial \sigma_e} = -\frac{1}{\sigma_e^2} \sum_{j=1}^l z_j Q_j, \quad (17)$$

$$\frac{\partial^2 \ln L_e(\theta_e)}{\partial \sigma_e \partial \mu_e} = -\frac{1}{\sigma_e^2} \left(\sum_{j=1}^l z_j Q_j \right.$$

$$\left. + \sum_{j=1}^l [(N_j - n_j) h_{e1}(z_j) - n_j h_{e2}(z_j)] \right), \quad (18)$$

$$\frac{\partial^2 \ln L_e(\theta_e)}{\partial \sigma_e^2} = -\frac{1}{\sigma_e^2} \left(\sum_{j=1}^l z_j^2 Q_j \right.$$

$$\left. + \sum_{j=1}^l z_j [(N_j - n_j) h_{e1}(z_j) - n_j h_{e2}(z_j)] \right). \quad (19)$$

Then

$$\mathbf{H}(\theta_e = \hat{\theta}_e) = \begin{bmatrix} -\frac{1}{\hat{\sigma}_e^2} \sum_{j=1}^l Q_j & -\frac{1}{\hat{\sigma}_e^2} \sum_{j=1}^l z_j Q_j \\ -\frac{1}{\hat{\sigma}_e^2} \sum_{j=1}^l z_j Q_j & -\frac{1}{\hat{\sigma}_e^2} \sum_{j=1}^l z_j^2 Q_j \end{bmatrix} \quad (20)$$

and

$$\mathbf{u}' \mathbf{H}(\theta_e = \hat{\theta}_e) \mathbf{u} = -\frac{1}{\hat{\sigma}_e^2} \sum_{j=1}^l (1 + z_j)^2 Q_j < 0, \quad (21)$$

i.e., the MLEs, $\hat{\mu}_e$ and $\hat{\sigma}_e$, exist and are unique ones that maximize (1). Thus, the following theorem has been proven.

Theorem 1. Suppose that X is a random variable coming from a situation described by a location-scale family of probability distribution functions, indexed by a parametric vector $\theta_e = (\mu_e, \sigma_e)$,

$$F_e(x | \theta_e) = \Pr(X \leq x)$$

$$= F_e\left(\frac{x - \mu_e}{\sigma_e}\right) = F_e(z), \quad -\infty < x < \infty, \quad (22)$$

where $-\infty < \mu_e < \infty$ is a location parameter and $\sigma_e > 0$ is a scale parameter, the distribution of $Z = (X - \mu_e) / \sigma_e$ does not depend on any unknown parameters.

To find an estimate $\hat{\theta}_e = (\hat{\mu}_e, \hat{\sigma}_e)$ of the unknown parametric vector $\theta_e = (\mu_e, \sigma_e)$, the likelihood function,

$$L_e(\theta_e) = \prod_{j=1}^l \binom{N_j}{n_j} [\Pr(X \leq d_j)]^{n_j} [\Pr(X > d_j)]^{N_j - n_j}$$

$$= \prod_{j=1}^l \binom{N_j}{n_j} F_e^{n_j}(d_j | \theta_e) [1 - F_e(d_j | \theta_e)]^{N_j - n_j}, \quad (23)$$

of the observed grouped data (n_1, \dots, n_l) and $(N_1 - n_1, \dots, N_l - n_l)$ from l random samples of sizes $N_j, j=1(1)l$, respectively, is used, where

$$0 < d_1 < d_2 < \dots < d_l \quad (24)$$

are fixed values,

$$0 \leq n_j \leq N_j, \forall j = 1(1)l. \quad (25)$$

Then the estimates, $\hat{\mu}_e$ and $\hat{\sigma}_e$, which maximize (23), exist and are unique if and only if the following conditions are satisfied:

- 1) $F_e(z)$ is strictly increasing continuous function for all $z \in (-\infty, \infty)$,
- 2) $F_e'(z)$ and $F_e''(z)$ exist and are continuous functions for all $z \in (-\infty, \infty)$,
- 3) $h_{e1}(z)$ is strictly increasing function for all $z \in (-\infty, \infty)$; $h_{e2}(z)$ is strictly decreasing function for all $z \in (-\infty, \infty)$.
- 4) $\sum_{j=1}^l N_j \sum_{j=1}^l d_j n_j > \sum_{j=1}^l n_j \sum_{j=1}^l d_j N_j$.

2.4. Estimation of Minimal Effective Dose

Numerical methods can be used to find the ML estimate $\hat{\theta}_e$. Then an estimate, d_{MED} , of the minimal effective dose (MED) for randomly chosen subjects, which elicits a prescribed lowest therapeutic response, is given by

$$d_{MED} = \arg[\Pr(X \leq d_{MED}) = 1 - \alpha], \quad (26)$$

where

$$\Pr(X \leq d_{MED}) = F_e(d_{MED} | \hat{\theta}_e). \quad (27)$$

α is a significance level (say, $\alpha = 0.05$).

3. Maximal Tolerated Dose Estimation

3.1. Likelihood Function of Grouped Data

Let us assume that the random variable Y , which represents a tolerated dose level of a drug (from the point of view of toxicity) for randomly chosen subject, has a continuous cumulative distribution function $F_t(y | \theta_t)$ (probability density function $f_t(y | \theta_t)$) with unknown parametric vector θ_t . We consider l ($l \geq$ the number of components of the unknown parametric vector θ_t) random samples of subjects of sizes $N_j, j=1(1)l$. The N_j subjects of the j th random sample are assigned to the dose d_j of a drug. Let m_j be the number of subjects in the j th sample, for which the tolerated dose of a drug is more than d_j . It is assumed, without loss of generality,

that $0 < d_1 < d_2 < \dots < d_l$.

The problem is to estimate the unknown parametric vector θ_t . For this purpose, it can be used the likelihood function of the grouped data

$$L_t(\theta_t) = \prod_{j=1}^l \binom{N_j}{m_j} [\Pr(Y \leq d_j)]^{N_j - m_j} [\Pr(Y > d_j)]^{m_j} \\ = \prod_{j=1}^l \binom{N_j}{m_j} F_t^{N_j - m_j}(d_j | \theta_t) [1 - F_t(d_j | \theta_t)]^{m_j}. \quad (28)$$

Consider a situation described by a location-scale family of probability distribution functions, indexed by the parametric vector $\theta_t = (\mu_t, \sigma_t)$,

$$F_t(y | \theta_t) = \Pr(Y \leq y) = F_t\left(\frac{y - \mu_t}{\sigma_t}\right) = F_t(z), \quad -\infty < y < \infty, \quad (29)$$

where $-\infty < \mu_t < \infty$ is a location parameter and $\sigma_t > 0$ is a scale parameter, the distribution of $Z = (Y - \mu_t) / \sigma_t$ does not depend on any unknown parameters.

Assumption 1. $F_t(z)$ is strictly increasing continuous function for all $z \in (-\infty, \infty)$, i.e., $F_t(z_2) > F_t(z_1)$ for all $z_2 > z_1$.

For computational convenience, the maximum likelihood estimate is obtained by maximizing the log-likelihood function, $\ln L_t(\theta_t)$. This is because the two functions, $L_t(\theta_t)$ and $\ln L_t(\theta_t)$, are monotonically related to each other so the same maximum likelihood estimate is obtained by maximizing either one. Thus,

$$\ln L_t(\theta_t) = \sum_{j=1}^l \ln \binom{N_j}{m_j} + \sum_{j=1}^l (N_j - m_j) \ln F_t(d_j | \theta_t) \\ + \sum_{j=1}^l m_j \ln [1 - F_t(d_j | \theta_t)] \\ = \sum_{j=1}^l \ln \binom{N_j}{m_j} + \sum_{j=1}^l (N_j - m_j) \ln F_t(z_j) \\ + \sum_{j=1}^l m_j \ln [1 - F_t(z_j)], \quad (30)$$

where

$$z_j = \frac{d_j - \mu_t}{\sigma_t}. \quad (31)$$

3.2. Log-Likelihood Equations

To obtain the log-likelihood equations for the unknown parameters μ_t and σ_t , the log-likelihood function in Equation (30) will be differentiated partially with respect to μ_t and σ_t , respectively, and set equal to zero to provide the MLEs, $\hat{\mu}_t$ and $\hat{\sigma}_t$, of the unknown parameters, μ_t and σ_t , respectively, as follows:

$$\frac{\partial \ln L_t(\theta_t)}{\partial \mu_t} = \frac{1}{\sigma_t} \sum_{j=1}^I [m_j h_{t1}(z_j) - (N_j - m_j) h_{t2}(z_j)] = 0, \quad (32)$$

$$\frac{\partial \ln L_t(\theta_t)}{\partial \sigma_t} = \frac{1}{\sigma_t} \sum_{j=1}^I z_j [m_j h_{t1}(z_j) - (N_j - m_j) h_{t2}(z_j)] = 0, \quad (33)$$

where

$$h_{t1}(z_j) = \frac{F_t'(z_j)}{1 - F_t(z_j)}, \quad h_{t2}(z_j) = \frac{F_t'(z_j)}{F_t(z_j)}. \quad (34)$$

Equations (32) and (33), which represent a necessary condition for the existence of the MLEs, can be reduced to:

$$\sum_{j=1}^I m_j h_{t1}(z_j) - \sum_{j=1}^I (N_j - m_j) h_{t2}(z_j) = 0, \quad (35)$$

$$\sum_{j=1}^I d_j m_j h_{t1}(z_j) - \sum_{j=1}^I d_j (N_j - m_j) h_{t2}(z_j) = 0. \quad (36)$$

It is usually not possible to obtain an analytic form solution of Equations (35) and (36). Hence, a numerical iterative technique can be used to obtain the MLEs, $\hat{\mu}_t$ and $\hat{\sigma}_t$, of the parameters, μ_t and σ_t , respectively.

3.3. Existence and Uniqueness of the ML Estimates

The maximum likelihood estimates need not exist nor be unique. In this section, we show that the MLEs, $\hat{\mu}_t$ and $\hat{\sigma}_t$, which maximize (28), exist in the above case and are unique.

Assumption 2. $F_t'(z)$ and $F_t''(z)$ exist and are continuous functions for all $z \in (-\infty, \infty)$.

Assumption 3. $h_{t1}(z)$ is strictly increasing function for all $z \in (-\infty, \infty)$; $h_{t2}(z)$ is strictly decreasing function for all $z \in (-\infty, \infty)$.

It follows from (35) and (36) that the MLEs, $\hat{\mu}_t$ and $\hat{\sigma}_t$, exist and are unique if and only if

$$\sum_{j=1}^I m_j \sum_{j=1}^I d_j (N_j - m_j) > \sum_{j=1}^I d_j m_j \sum_{j=1}^I (N_j - m_j) \quad (37)$$

or

$$\frac{\sum_{j=1}^I d_j m_j}{\sum_{j=1}^I d_j N_j} < \frac{\sum_{j=1}^I m_j}{\sum_{j=1}^I N_j}. \quad (38)$$

A sufficient condition for $\hat{\theta}_t = (\hat{\mu}_t, \hat{\sigma}_t)$ to be a maximum of (30) is that the Hessian matrix

$$\mathbf{H}(\theta_t) = \begin{bmatrix} \frac{\partial^2 \ln L_t(\theta_t)}{\partial \mu_t^2} & \frac{\partial^2 \ln L_t(\theta_t)}{\partial \mu_t \partial \sigma_t} \\ \frac{\partial^2 \ln L_t(\theta_t)}{\partial \sigma_t \partial \mu_t} & \frac{\partial^2 \ln L_t(\theta_t)}{\partial \sigma_t^2} \end{bmatrix} \quad (39)$$

evaluated at $\theta_t = \hat{\theta}_t$ satisfies the following condition [11]:

$$\mathbf{u}' \mathbf{H}(\theta_t) \mathbf{u} < 0, \quad (40)$$

where

$$\mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \quad (41)$$

It follows from (32) and (33) that

$$\frac{\partial^2 \ln L_t(\theta_t)}{\partial \mu_t^2} = -\frac{1}{\sigma_t^2} \sum_{j=1}^I Q_j, \quad (42)$$

where

$$Q_j = m_j h_{t1}'(z_j) - (N_j - m_j) h_{t2}'(z_j) > 0, \quad (43)$$

$$\frac{\partial^2 \ln L_t(\theta_t)}{\partial \mu_t \partial \sigma_t} = -\frac{1}{\sigma_t^2} \sum_{j=1}^I z_j Q_j, \quad (44)$$

$$\frac{\partial^2 \ln L_t(\theta_t)}{\partial \sigma_t \partial \mu_t} = -\frac{1}{\sigma_t^2} \left(\sum_{j=1}^I z_j Q_j \right.$$

$$\left. + \sum_{j=1}^I [m_j h_{t1}(z_j) - (N_j - m_j) h_{t2}(z_j)] \right), \quad (45)$$

$$\frac{\partial^2 \ln L_t(\theta_t)}{\partial \sigma_t^2} = -\frac{1}{\sigma_t^2} \left(\sum_{j=1}^I z_j^2 Q_j \right.$$

$$\left. + \sum_{j=1}^I z_j [m_j h_{t1}(z_j) - (N_j - m_j) h_{t2}(z_j)] \right). \quad (46)$$

Then

$$\mathbf{H}(\theta_t = \hat{\theta}_t) = \begin{bmatrix} -\frac{1}{\hat{\sigma}_t^2} \sum_{j=1}^I Q_j & -\frac{1}{\hat{\sigma}_t^2} \sum_{j=1}^I z_j Q_j \\ -\frac{1}{\hat{\sigma}_t^2} \sum_{j=1}^I z_j Q_j & -\frac{1}{\hat{\sigma}_t^2} \sum_{j=1}^I z_j^2 Q_j \end{bmatrix} \quad (47)$$

and

$$\mathbf{u}' \mathbf{H}(\theta_t = \hat{\theta}_t) \mathbf{u} = -\frac{1}{\hat{\sigma}_t^2} \sum_{j=1}^I (1 + z_j)^2 Q_j < 0. \quad (48)$$

i.e., the MLEs, $\hat{\mu}_e$ and $\hat{\sigma}_e$, exist and are unique ones that maximize (28). Thus, the following theorem has been proven.

Theorem 2. Suppose that Y is a random variable coming from a situation described by a location-scale family of probability distribution functions, indexed by the parametric vector $\theta_t = (\mu_t, \sigma_t)$,

$$F_t(y | \theta_t) = \Pr(Y \leq y)$$

$$= F_t \left(\frac{y - \mu_t}{\sigma_t} \right) = F_t(z), \quad -\infty < y < \infty, \quad (49)$$

where $-\infty < \mu_t < \infty$ is a location parameter and $\sigma_t > 0$ is a scale parameter, the distribution of $Z = (Y - \mu_t) / \sigma_t$ does not depend on any unknown parameters.

To find an estimate $\hat{\theta}_t = (\hat{\mu}_t, \hat{\sigma}_t)$ of the unknown parametric vector $\theta_t = (\mu_t, \sigma_t)$, the likelihood function,

$$L_t(\theta_t) = \prod_{j=1}^l \binom{N_j}{m_j} [\Pr(Y \leq d_j)]^{N_j - m_j} [\Pr(Y > d_j)]^{m_j} \\ = \prod_{j=1}^l \binom{N_j}{m_j} F_t^{N_j - m_j}(d_j | \theta_t) [1 - F_t(d_j | \theta_t)]^{m_j}. \quad (50)$$

of the observed grouped data (m_1, \dots, m_l) and $(N_1 - m_1, \dots, N_l - m_l)$ from l random samples of sizes $N_j, j=1(1)l$, respectively, is used, where

$$0 < d_1 < d_2 < \dots < d_l, \quad (51)$$

are fixed values,

$$0 \leq m_j \leq N_j, \forall j = 1(1)l. \quad (52)$$

Then the estimates, $\hat{\mu}_t$ and $\hat{\sigma}_t$, which maximize (50), exist and are unique if and only if the following conditions are satisfied:

- 1) $F_t(z)$ is strictly increasing continuous function for all $z \in (-\infty, \infty)$,
- 2) $F_t'(z)$ and $F_t''(z)$ exist and are continuous functions for all $z \in (-\infty, \infty)$,
- 3) $h_{t1}(z)$ is strictly increasing function for all $z \in (-\infty, \infty)$; $h_{t2}(z)$ is strictly decreasing function for all $z \in (-\infty, \infty)$.
- 4) $\sum_{j=1}^l N_j \sum_{j=1}^l d_j m_j < \sum_{j=1}^l m_j \sum_{j=1}^l d_j N_j$.

3.4. Estimation of Maximal Tolerated Dose

Numerical methods can be used to find the ML estimate $\hat{\theta}_t$. Then an estimate, d_{MTD} , of a safe dose (maximal tolerated dose (MTD)) for randomly chosen subjects that will not cause some undesirable effect (e.g., toxicity, carcinogenicity) is given by

$$d_{\text{MTD}} = \arg [\Pr(Y > d_{\text{MTD}}) = 1 - \alpha], \quad (53)$$

where

$$\Pr(Y > d_{\text{MTD}}) = 1 - F_t(d_{\text{MTD}} | \hat{\theta}_t) = \bar{F}_t(d_{\text{MTD}} | \hat{\theta}_t), \quad (54)$$

α is a significance level (say, $\alpha = 0.05$).

It is clear that upper end of the interval is the maximal

tolerated dose (MTD) d_{MTD} , and the lower end of the interval is the minimal effective dose (MED) d_{MED} , i.e., $d_{\text{MED}} < d_{\text{MTD}}$.

Statistical inference. If the dose of a drug d is given by

$$d \in (d_{\text{MED}}, d_{\text{MTD}}), \quad (55)$$

then there is more than $100(1 - \alpha)\%$ assurance that d will be a suitable dose for randomly chosen subjects from the point of view of efficacy as well as from the point of view of toxicity.

4. Numerical Examples

4.1. Example 1

Let us assume that the random variable X , which represents an effective dose level of a drug (from the point of view of efficacy) for randomly chosen subject, has a continuous cumulative distribution function of a normal distribution, i.e.,

$$X \sim F_e(x | \theta_e) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x - \mu_e) / \sigma_e} \exp\left(-\frac{z^2}{2}\right) dz, \quad (56)$$

where $\theta_t = (\mu_t, \sigma_t)$ is an unknown parametric vector, $-\infty < \mu_t < \infty$ is a location parameter and $\sigma_t > 0$ is a scale parameter. There are $l=4$ random samples of subjects of size $N_j=10, j=1(1)l$, and $l=4$ fixed doses $d_j, j=1(1)l$, of a drug, where

$$0 < (d_1 = 2) < (d_2 = 6) < (d_3 = 10) < (d_4 = 14). \quad (57)$$

The $N_j=10$ subjects of the j th random sample are assigned to the dose d_j of a drug. Let n_j be the number of subjects in the j th sample, for which the effective dose of a drug is less than d_j .

The observed grouped data n_j and $N_j - n_j$ from l random samples of sizes $N_j, j=1(1)l$, respectively, are given as follows:

$$n_1 = 1, \quad n_2 = 9, \quad n_3 = 10, \quad n_4 = 10 \quad (58)$$

and

$$N_1 - n_1 = 9, \quad N_2 - n_2 = 1, \quad N_3 - n_3 = 0, \quad N_4 - n_4 = 0. \quad (59)$$

It follows from (11) that

$$\frac{\sum_{j=1}^l d_j n_j}{\sum_{j=1}^l d_j N_j} = 0.925 > \frac{\sum_{j=1}^l n_j}{\sum_{j=1}^l N_j} = 0.75, \quad (60)$$

i.e., the MLEs, $\hat{\mu}_e$ and $\hat{\sigma}_e$, exist and are unique ones that maximize (1). Thus, using the Solver of Excel 2010 for maximizing (1), we have that

$$\hat{\theta}_e = \arg \left[\max_{\theta_e} L_e(\theta_e) \right] = (\hat{\mu}_e, \hat{\sigma}_e), \quad (61)$$

where

$$\hat{\mu}_e = 3.999446, \quad \hat{\sigma}_e = 1.559312336. \quad (62)$$

$$\max_{\theta_e} L_e(\theta_e) = L_e(\hat{\theta}_e) = 0.0015. \quad (63)$$

Then the estimate, d_{MED} , of the minimal effective dose (MED) for randomly chosen subjects, which elicits a prescribed lowest therapeutic response, is given by

$$\begin{aligned} d_{\text{MED}} &= \arg [\Pr(X \leq d_{\text{MED}}) = 1 - \alpha] \\ &= \arg \left[F_e(d_{\text{MED}} | \hat{\theta}_e) = 1 - \alpha \right] = 7.0557, \end{aligned} \quad (64)$$

where the significance level $\alpha = 0.025$.

Let us assume that the random variable Y , which represents a tolerated dose level of a drug (from the point of view of toxicity) for randomly selected subject, also has a continuous cumulative distribution function of a normal distribution, i.e.,

$$Y \sim F_t(x | \theta_t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(y-\mu_t)/\sigma_t} \exp\left(-\frac{z^2}{2}\right) dz. \quad (65)$$

The observed grouped data m_j and $N_j - m_j$ from l random samples of sizes $N_j, j=1(1)l$, respectively, are given as follows:

$$m_1 = 10, \quad m_2 = 10, \quad m_3 = 9, \quad m_4 = 1 \quad (66)$$

and

$$N_1 - m_1 = 0, \quad N_2 - m_2 = 0, \quad N_3 - m_3 = 1, \quad N_4 - m_4 = 9. \quad (67)$$

It follows from (38) that

$$\frac{\sum_{j=1}^l d_j m_j}{\sum_{j=1}^l d_j N_j} = 0.575 < \frac{\sum_{j=1}^l m_j}{\sum_{j=1}^l N_j} = 0.75. \quad (68)$$

i.e., the MLEs, $\hat{\mu}_e$ and $\hat{\sigma}_e$, exist and are unique ones that maximize (28). Thus, using the Solver of Excel 2010 for maximizing (28), we have that

$$\hat{\theta}_t = \arg \left[\max_{\theta_t} L_t(\theta_t) \right] = (\hat{\mu}_t, \hat{\sigma}_t), \quad (69)$$

where

$$\hat{\mu}_t = 12.00055404, \quad \hat{\sigma}_t = 1.559312; \quad (70)$$

$$\max_{\theta_t} L_t(\theta_t) = L_t(\hat{\theta}_t) = 0.001500047. \quad (71)$$

Then the estimate, d_{MTD} , of a safe dose (maximal tolerated dose (MTD)) for randomly chosen subject, which will not cause some undesirable effect (e.g., toxicity, carcinogenicity) is given by

$$\begin{aligned} d_{\text{MTD}} &= \arg [\Pr(Y > d_{\text{MTD}}) = 1 - \alpha] \\ &= \arg \left[\bar{F}_t(d_{\text{MTD}} | \hat{\theta}_t) = 1 - \alpha \right] = 8.9445, \end{aligned} \quad (72)$$

where the significance level $\alpha = 0.025$.

Thus, the upper end of the interval is the maximal tolerated dose (MTD) $d_{\text{MTD}} = 8.9445$ and the lower end of the interval is the minimal effective dose (MED) $d_{\text{MED}} = 7.0557$, i.e., $d_{\text{MED}} = 7.0557 < d_{\text{MTD}} = 8.9445$.

Statistical inference. If the dose d of a drug is given by

$$d \in (d_{\text{MED}} = 7.0557, d_{\text{MTD}} = 8.9445), \quad (73)$$

then there is more than 97.5% assurance that d will be a suitable dose for randomly chosen subjects from the point of view of efficacy as well as from the point of view of toxicity.

4.2. Example 2

Consider the situation described in Example 1. Let us assume that

$$X \sim F_{e(W)}(x | \theta_e), \quad (74)$$

where

$$F_{e(W)}(x | \theta_e) = \Pr(X \leq x) = 1 - \exp\left[-\left(\frac{x}{\beta_e}\right)^{\delta_e}\right], \quad x > 0, \quad (75)$$

is a cumulative distribution function of a two-parameter Weibull distribution with an unknown parametric vector $\theta_e = (\beta_e, \delta_e)$, $\beta_e > 0$ is a scale parameter, $\delta_e > 0$ is a shape parameter. Using the Solver of Excel 2010 for maximizing the likelihood function of the grouped data (58) and (59),

$$\begin{aligned} L_{e(W)}(\theta_e) &= \prod_{j=1}^l \binom{N_j}{n_j} [\Pr(X \leq d_j)]^{n_j} [\Pr(X > d_j)]^{N_j - n_j} \\ &= \prod_{j=1}^l \binom{N_j}{n_j} F_{e(W)}^{n_j}(d_j | \theta_e) [1 - F_{e(W)}(d_j | \theta_e)]^{N_j - n_j}, \end{aligned} \quad (76)$$

it can be obtained the MLEs

$$\hat{\theta}_e = \arg \left[\max_{\theta_e} L_{e(W)}(\theta_e) \right] = (\hat{\beta}_e, \hat{\delta}_e), \quad (77)$$

where

$$\hat{\beta}_e = 4.457434, \hat{\delta}_e = 2.811425; \tag{78}$$

$$\max_{\theta_e} L_{e(W)}(\theta_e) = L_{e(W)}(\hat{\theta}_e) = 0.0015. \tag{79}$$

Then the estimate, d_{MED} , of the minimal effective dose (MED) for randomly selected subjects, which elicits a prescribed lowest therapeutic response, is given by

$$\begin{aligned} d_{MED} &= \arg [\Pr(X \leq d_{MED}) = 1 - \alpha] \\ &= \arg [F_{e(W)}(d_{MED} | \hat{\theta}_e) = 1 - \alpha] = 7.0913, \end{aligned} \tag{80}$$

where the significance level $\alpha = 0.025$.

Now, let us assume that

$$Y \sim F_{t(W)}(y | \theta_t), \tag{81}$$

where

$$F_{t(W)}(y | \theta_t) = \Pr(Y \leq y) = 1 - \exp \left[- \left(\frac{y}{\beta_t} \right)^{\delta_t} \right], \quad y > 0, \tag{82}$$

is a cumulative distribution function of a two-parameter Weibull distribution with an unknown parametric vector $\theta_t = (\beta_t, \delta_t)$, $\beta_t > 0$ is a scale parameter, $\delta_t > 0$ is a shape parameter. Using the Solver of Excel 2010 for maximizing the likelihood function of the grouped data (66) and (67),

$$\begin{aligned} L_{t(W)}(\theta_t) &= \prod_{j=1}^l \binom{N_j}{m_j} [\Pr(Y \leq d_j)]^{N_j - m_j} [\Pr(Y > d_j)]^{m_j} \\ &= \prod_{j=1}^l \binom{N_j}{m_j} F_{t(W)}^{N_j - m_j}(d_j | \theta_t) [1 - F_{t(W)}(d_j | \theta_t)]^{m_j}, \end{aligned} \tag{83}$$

it can be obtained the MLEs

$$\hat{\theta}_t = \arg \left[\max_{\theta_t} L_{t(W)}(\theta_t) \right] = (\hat{\beta}_t, \hat{\delta}_t), \tag{84}$$

where

$$\hat{\beta}_t = 12.78881699, \hat{\delta}_t = 9.243318; \tag{85}$$

$$\max_{\theta_t} L_{t(W)}(\theta_t) = L_{t(W)}(\hat{\theta}_t) = 0.001487. \tag{86}$$

Then the estimate, d_{MTD} , of a safe dose (maximal tolerated dose (MTD)) for randomly chosen subject, which will not cause some undesirable effect (e.g., toxicity, carcinogenicity) is given by

$$\begin{aligned} d_{MTD} &= \arg [\Pr(Y > d_{MTD}) = 1 - \alpha] \\ &= \arg \left[\bar{F}_{t(W)}(d_{MTD} | \hat{\theta}_t) = 1 - \alpha \right] = 8.59218, \end{aligned} \tag{87}$$

where the significance level $\alpha = 0.025$.

Maximum likelihood inference. Since

$$L_e(\hat{\theta}_e) = 0.0015 = L_{e(W)}(\hat{\theta}_e) = 0.0015 \tag{88}$$

and

$$L_t(\hat{\theta}_t) = 0.001500047 > L_{t(W)}(\hat{\theta}_t) = 0.001487, \tag{89}$$

it follows from (88) and (89) that the normal distribution better fits the grouped data of Example 1 than the Weibull distribution.

5. Conclusion

Dose–response experiments are an important part of biomedical research to study relationships between increasing doses of a therapeutic compound and a variety of responses. Typically, the response represents a phenotypical effect of a compound such as inhibition, stimulation, toxicity, or expression level of a certain gene. The primary goal of such an experiment is to detect a dose–response relationship and to determine the nature of the relationship wherever it exists. In this article, we focus on a continuous response and an experimental design with fixed number of doses. If the dose–response relationship exists, it may be monotone, that is, the compound effect (increasing or decreasing) becomes stronger (or stays the same) with an increasing dose. This property is very common in real applications, especially when inhibition or toxicity is measured. More general umbrella-shaped profiles [12] can occur within the context of overdosing and therefore a decreasing (increasing) effect is expected after reaching some threshold dose. These properties are not discussed in this paper. It is assumed only that the dose of a drug suitable for a randomly chosen subject is a random variable that follows a certain law of probability distribution. If there are a few models of the probability distribution for the aforementioned random variable, then the best model is the one that provides the greatest value of the maximum likelihood function of the grouped data. The methodology described here can be extended in several different directions to handle various problems that arise in practice. We have illustrated the proposed methodology for location-scale distributions (such as, say, the normal, lognormal, Gumbel, Weibull distribution, etc.). Applications to other distributions could follow directly.

References

- [1] D.T. Greenwood and A. H. Todd, “From laboratory to clinical use,” in *Clinical Trials*, Johnson, F. N. and Johnson, S. (eds), Blackwell, London, 1977, pp. 50–70.
- [2] J. R. Crout and M. J. Finkel, *General Considerations for the Clinical Evaluation of 4 Drugs*. U.S. Dept. of Health, Education and Welfare-Public Health Service F. D. A., 1977.

- [3] H. Robbins and S. Monro, "A stochastic approximation method," *Annals of Mathematical Statistics*, vol. 22, pp. 400 – 407, 1951.
- [4] C.-F.J. Wu, "Efficient sequential designs with binary data," *J. Amer. Statist. Assoc.*, vol 80, pp. 974 – 984, 1985.
- [5] B. H. Eichhorn and S. Zacks, "Sequential search of an optimal dosage," *J. Amer. Statist. Assoc.*, vol. 68, pp. 594 –598, 1973.
- [6] B. H. Eichhorn and S. Zacks, "Bayes sequential search of an optimal dosage: linear regression with both parameters unknown," *Communications in Statistics-Theory and Methods*, vol. 10, pp. 931– 953, 1981.
- [7] J. Pinheiro, F. Bretz, and M. Branson, *Dose Finding in Drug Development*. New York: Springer, 2006.
- [8] B. Bornkamp, J. C. Pinheiro, and F. Bretz, "MCPMod: An R package for the design and analysis of dose-finding studies," *Journal of Statistical Software*, vol. 29, pp. 1–23, 2009.
- [9] M. Whitney and L. Ryan, "Quantifying dose-response uncertainty using Bayesian model averaging," in *Uncertainty Modeling in Dose Response: Bench Testing Environmental Toxicity*. New York: Wiley, 2009, pp. 165–179.
- [10] M. Gasparini and J. Eisele, "A curve-free method for phase I clinical trials," *Biometrics*, vol. 56, pp. 609 – 615, 2000.
- [11] I. J. Myung, "Tutorial on maximum likelihood estimation," *Journal of Mathematical Psychology*, vol. 47, pp. 90–100, 2003.
- [12] F. Bretz and L. A. Hothorn, "Statistical analysis of monotone or non-monotone dose-response data from *in vitro* toxicological assays," *Alternatives to Lab Animals*, vol. 31, pp.81– 96, 2003.