7-6-2004

On Effective and Efficient Experimental Designs for Neurobehavioral Screening Tests: The Choice of a Testing Time for Estimating the Time of Peak Effects

Peter A. Toyinbo
University of South Florida

On Effective and Efficient Experimental Designs for Neurobehavioral Screening Tests:

The Choice of a Testing Time for Estimating the Time of Peak Effects

by

Peter A. Toyinbo

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Public Health
Department of Epidemiology and Biostatistics
College of Public Health
University of South Florida

Major Professor: Yiliang Zhu, Ph.D.
Member: C. Hendricks Brown, Ph.D.
Member: Getachew Dagne, Ph.D.

Date of Approval:
July 06, 2004

Keywords: Neurotoxicity, dose response, modeling, methodology, benchmark dose.

DEDICATION

To Grace

For as long as we can remember, each step of our remarkable journey, you continue to be an inspiration to us. You freely bestow to our family all that you have, all that you are. Gracefully, you led us to one conviction: we are blessed with a lasting amazing Grace.

We love you, Grace.

Peter, for Funbi, Femi and Tomi.

ACKNOWLEDGMENTS

I am grateful to Dr. Yiliang Zhu, my advisor and major professor, for his encouragement, support, research expertise, critical evaluation and excellent guidance in grooming this thesis into its end product. In addition, the gracious time, technical guidance and professional advice that Dr. Wei Wang gave were invaluable to the contents of this thesis.

I gratefully acknowledge Dr. Hendricks Brown and Dr. Getachew Dagne for their participation on this thesis committee and for sharing key insights and valued expertise in the completion of the thesis.

A special thank to my friends and colleagues in the Health Risk Assessment Methodology Group of the University of South Florida College of Public Health: Michael Wessel, Jenny Jia and Dr. Sanjay Muniredi for all the hours of brain storming, thoughtful comments, and relaxing fun that went into this thesis.

I wish to acknowledge and thank Dr. Virginia Moser of US EPA as the provider of the FOB data used to carry out this thesis.

Finally, as a fruitful tree would acknowledge its root for its nurture, I am incredibly indebted to my family for their support, patience, understanding, and for believing in me. Especially to Grace, my wife, this accomplishment is the resultant of your countless gentle nudges.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ON EFFECTIVE AND EFFICIENT EXPERIMENTAL DESIGNS FOR NEUROBEHAVIORAL SCREENING TESTS: THE CHOICE OF A TESTING TIME FOR ESTIMATING THE TIME OF PEAK EFFECTS

Peter A. Toyinbo

ABSTRACT

In its latest neurotoxicity guidelines released by the US EPA Office of Prevention, Pesticides and Toxic Substances (OPPTS) in 1998, it is recommended that in a neurobehavioral testing, at a minimum, for acute studies, observations and activity testing should be made before the initiation of exposure, at the estimated TOPE (time of peak effects) within 8 hrs of dosing, and at 7 and 14 days after dosing. It is recommended that estimation of TOPE be made by dosing pairs of rats across a range of doses and making regular observations of gait and arousal. However it is well known that TOPE may vary with end points or exposure conditions.

In order to derive quantitative safety measures such as the benchmark doses (BMD), dose-time-response modeling must be done first and a model-based estimate is then implied. In many cases, the overall BMD corresponds to a TOPE estimate. In such cases a substantial variation in the TOPE estimate in turn may result in substantial variation in BMD estimate. Therefore a reliable statistical estimate of TOPE is crucial to the correct determination of BMD.

We therefore performed simulation studies to assess the impact of the experiment-based TOPE on the statistical estimation of the true TOPE on the basis of a fitted dose-

time-response model. The simulation allows for the determination of the optimal timing range for the 2nd testing.

The results indicated that given only four repeated observations, the optimal second testing time was at about midway between time zero and the true TOPE. Choosing the second testing time at the TOPE may not generate statistical estimates closer to the true TOPE.

CHAPTER 1

INTRODUCTION

1.1 Neurotoxicity and Neurobehavioral Screening Methods

Chemicals are an integral part of life, with the capacity to improve as well as endanger health. A link between human exposure to some chemical substances and neurotoxicity has been firmly established (Anger, 1986; US EPA, 1990). Neurotoxicity is defined as adverse effects on either the structure or functions of the nervous system (US EPA, 1998a). In addition to its primary role in psychological functions, the nervous system controls most, if not all, other bodily processes. Nervous system is sensitive to perturbation from various sources and has limited ability to regenerate. Therefore, there is a need for regulation of neurotoxicants on a scientific basis. It is important to have consistent guidance on how to evaluate neurotoxic substances and assess their potential to cause transient or persistent, direct or indirect effects on human health. (US EPA, 1998a)

In the EPA's neurotoxicity risk assessment guidelines (US EPA, 1998a), five categories of endpoints were described: structural or neuropathological, neurophysiological, neurochemical, behavioral, and developmental. The guidelines outline the scientific basis for evaluating effects due to exposure to neurotoxicants and discuss principles and methods for evaluating data from human and animal studies using the described endpoints.

In collaboration with international organizations such as the International Programme on Chemical Safety (IPCS), the US EPA has been developing and evaluating test methods that may eventually lead to an integrated approach to risk assessment of neurotoxicity. The EPA recommended the use of neurobehavioral screening methods as a first tier test for identifying and quantifying neurotoxicity of chemicals from animal studies (MacPhail et al, 1997; US EPA, 1998a). One such neurotoxicity screening battery is the Functional Observation Battery (FOB) in conjunction with motor activity (US EPA, 1998b). For the purpose of this thesis, FOB will be construed to encompass neurobehavioral screening methods.

1.2 Time of Peak Effects and Benchmark Dose Estimation

There have been increasing efforts to improving the scientific methodologies for risk assessment of neurotoxic effects in human due to chemical exposure. US EPA has recommended to use BMD as an alternative to the NOAEL/LOAEL methodology (US EPA, 1998a). The benchmark dose (BMD) approach aims to identify an effective dose (ED) that would induce an increase (typically 1-10%) of the attributable risk of adverse effects over background through empirical modeling (Crump, 1984; Zhu 2001). This approach provides for more quantitative dose-response evaluation when sufficient data are available and it takes into account the variability in the data and the slope of the dose-response curve. (Crump, 1984; U.S. EPA, 1995; Zhu, 2001).

A number of non-linear mixed effects models have been developed for describing the dose-time-response relationships observed in the FOB data from the EPA Superfund study and the IPCS Collaborative study (Zhu, 2001; Zhu et al, 2003a,b). Methods to

implement benchmark dose methodology for neurotoxicity data have also been developed (Zhu et al, 2003c). Zhu (2003) showed that both estimates of attributable risk and BMD vary with exposure level, time of testing, and spontaneous risk. He argues that a time profile of BMD be considered and the smallest value over the time course be reported as the overall value for deriving a safety dose in regulation. For some dose-response models (that will be focus of this thesis), this overall BMD must correspond to the time of peak effects (TOPE). A reliable estimate of TOPE is therefore crucial to the correct determination of BMD.

1.3 Neurobehavioral Screening Protocol

Neurotoxicity testing procedures must meet certain data requirements of the U.S. EPA under the Toxic Substances Control Act and the Federal Insecticide, Fungicide and Rodenticide Act (US EPA, 1991). In order to minimize variations among the testing procedures, the US EPA Office of Prevention, Pesticides and Toxic Substances (OPPTS) harmonized several other guidelines into a single set of OPPTS guidelines released in 1998 (US EPA, 1998b). Specifically for single dose experiments, the OPPTS guidelines include the following recommendation for time of testing: "At a minimum, for acute studies, observations and activity testing should be made before the initiation of exposure, at the estimated TOPE (time of peak effects) within 8 hrs of dosing, at 7 and 14 days after dosing. Estimation of TOPE may be made by dosing pairs of rats across a range of doses and making regular observations of gait and arousal."

The OPPTS guidelines of 1998 (US EPA, 1998b) was preceded by a similar design protocol that was adopted in the IPCS Collaborative study (Moser et al, 1997a,b)

and which produced the FOB data used in this thesis. Under the Collaborative study protocol, for the acute exposure experiments, FOB and motor activity measurements were conducted at four testing times:

t_1 immediately prior to exposure,

t_2 estimated time of peak effect (TOPE),

t_3 one day after dosing, and

t_4 seven days after dosing.

1.4 Experiment Based TOPE Estimate

The endpoints of FOB tests consist of about 30 non-invasive measures of gross functional deficits that quantify neurobehavioral changes in animals exposed to a chemical substance. The FOB measures can be grouped into six neurobehavioral functional domains, including activity, neuromuscular, excitability, sensorimotor, physiological and autonomic functions (McDaniel and Moser, 1993; Moser et al, 1997a).

Whereas individual endpoints can be used for risk assessment, there are efforts to explore the use of composite domain scores. Obviously it is practically inefficient to employ all available endpoints in a pilot study. Alternatively, the US EPA recommended that the method for selecting time of testing to be used in acute studies be based on range-finding pilot study using gait and arousal as the endpoints for determining TOPE (Moser et al, 1997a; US EPA, 1998b), thus reducing the number of endpoints to a more manageable size of two. As a result, the experimentally determined TOPE is by design unique to individual chemical agents. However a previous study has shown that when the recommended end-points for determining the TOPE in a pilot study are limited to only

gait and arousal, the second testing time (t_2) selected (assuming four testing times) for the acute study proper based on this TOPE estimate may not be appropriate for other neurotoxic effects or endpoints which show a different time course (Lammers and Kulig, 1997). Conceivably, apart from random error of measurement, the TOPE estimate thus obtained might systematically differ from the true TOPE (parameter). For this reason, the timing adopted for the second testing may differ from the true TOPE substantially even following the EPA guidelines. The impact of such selection is largely unknown.

1.5 Model Based TOPE Estimate

The FOB measures were multi-scale and also were grouped into six functional domains. In order to reduce the number of endpoints for statistical efficiency, these multi-scale measures were converted to domain-specific composite scores (McDaniel and Moser, 1993; Zhu et al, 2003b). Typically, a composite score would be a weighted average of individual scores involved. According to Zhu et al (2003b) this approach mandates, as a prerequisite, conversion of individual measures to a common ordinal scale. The authors therefore converted every measure, continuous or categorical, into a 4-level ordinal scale in which ranking of an observation was based on the extent to which the corresponding neurobehavioral response was "common" in occurrence in a reference group.

Zhu et al (2003b) then proceeded to dose-time-response modeling of the domain-specific composite scores, i.e. of grouped FOB measures, as part of the steps leading to BMD estimation. For each acute experiment (or chemical), a statistical model was fitted separately to each of six domains to produce a total of six domain-specific TOPE

estimates. It was these model-based TOPE estimates that were used to compute the benchmark dose for each composite score.

Expectedly, for individual chemicals, each of the six domain-specific model-based TOPE estimates might be different than the single experiment-based TOPE estimate used from the pilot study. While differences in values between these two types of TOPE estimates are expected, the reliability of model-based TOPE estimates cannot be presumed. It is conceivable that the reliability of the model-based TOPE estimates might also be affected by the uncertainty inherent in the timing of the 2nd testing that was determined from experiment-based TOPE estimate.

1.6 Objectives of this Study

We believe that any uncertainty about the TOPE derived from the pilot experiment is carried over to variability in the timing of the 2nd testing. Furthermore, it is not clear how variability in the 2nd testing time around the true TOPE could impact the statistical estimation of the true TOPE on the basis of a fitted dose-time-response model. Clearly there is a need for effective experimental designs to facilitate the estimation of the true underlying TOPE by any well fitted model. On the contrary, a poorly designed experiment often results not only in inefficient use of time and other resources, but also in invalid (bias) and/or imprecise (large variation) estimation. Therefore investigating effective and efficient time points in FOB tests for identifying TOPE may lead to improved neurotoxicity screening procedures. It is this aspect of the neurobehavioral screening protocol for acute experiments that this thesis will focus on.

The main research questions this thesis seeks to answer are as follows:

1. Under the proposed protocol of the US EPA/IPCS Collaborative study, what impacts would the timing of the 2nd testing have on estimating the true TOPE?

2. If we fix the 1st, 3rd and 4th FOB testing times, what would be the optimal range for the 2nd testing time to effectively estimate the TOPE?

3. How sensitive are the non-linear mixed effects models under consideration to variability in the 2nd testing times?

CHAPTER 2

THEORY AND METHODS

2.1 Dose-Time-Response Models

A family of three linear/nonlinear dose-response models with random effects (Zhu et al, 2003a,b.) were fitted to the Functional Observation Battery (FOB) and an automated motor activity data from the EPA/IPCS Collaborative Study (Moser et al, 1997b). The three statistical models are Linear-Exponential, Complementary-Exponential and Toxico-Diffusion models. The first two are different forms of the diffusion model as briefly described below. The diffusion model describes the expected response as a function of dose and time:

$$Expected\ \ response = f(t,d) = A + \frac{Btd\exp(-K_e t)}{1 + Ctd\exp(-K_e t)}$$

where t = testing time; d = administered dose, and B, C and K_e ("elimination rate") are parameters to be estimated. A is the baseline level and can be time dependent. If the coefficient C=0, we have the linear exponential model given by

$$Expected\ \ response = f(t,d) = A + Btd\exp(-K_e t)$$

Linearization of the diffusion model with respect to $Ctd\exp(-K_e t)$ via first order Taylor series expansion leads to the complementary exponential model:

$$Expected\ \ response = f(t,d) = A + Btd\exp(-K_e t)\{1 - Ctd\exp(-K_e t)\}$$

In all three models the "elimination rate" K_e plays an important role. As *t* varies from $[0,\infty)$, *f(t,d)* attains an extreme value (either maximum or minimum depending on the sign of B) at $t=1/K_e$, then returns towards the baseline *f(0,0)*. These three models are capable of modeling neurotoxic effects that are transient in time, with a common time of peak effects (TOPE) at $t=1/K_e$ irrespective of exposure level. Of the three, only the Linear-Exponential and Toxico-Diffusion were used as cases in this thesis.

Another non-linear model that has never been fitted to the FOB data was also used in this thesis. Unlike the three models previously described, this model is non-exponential and non-mechanistic in any sense. It is a simple rational function hence it is referred to as Rational Function model and is given by

$$response = f(t,d) = A + \frac{Btd}{K + t^2}$$

Here the TOPE is also independent of the exposure dose and is computed from estimable *K* parameter. As *t* varies from $[0,\infty)$, *f(t,d)* peaks to a maximum (*B*>0) at $t = \sqrt{K}$ (TOPE), then decreases back towards *A*. The inclusion of this non-exponential model would permit us to further examine the sensitivity of designs to the underlying models.

2.2 TOPE Estimation

Statistical modeling of a sample data such as the FOB data aims to capture and describe the underlying distribution of the data in an analytical way so that it is understandable and interpretable systematically. The TOPE, our parameter of interest,

must be estimated directly from a model fit to the data. The reliability of the TOPE estimate depends upon both the statistical estimator and the experimental design. Ideally we would like expected value of the estimator to equal the parameter estimated; that is $E(\hat{\theta}) = \theta$, where θ is the population parameter and $\hat{\theta}$ is the point estimator of θ. The point estimator is said to be unbiased if the bias $B = E(\hat{\theta}) - \theta = 0$. In addition we would also prefer that the variation of the estimator $V(\hat{\theta})$ be as small as possible because a smaller variance indicates that under replications, a higher fraction of values of $\hat{\theta}$ will be "close" to θ. The overall accuracy of the point estimator $\hat{\theta}$ can be characterized by the mean squared error (MSE) that combines variance and bias to form a single measure.

$$MSE(\hat{\theta}) = V(\hat{\theta}) + B^2$$

Thus, assuming we know the true population dose-time-response trend, we can numerically measure the overall quality of a statistical estimator of the TOPE by computing both the bias and the mean squared error. However a statistical estimator and its properties generally depend on the experimental design that generates the data. In the FOB tests, for example, the different sets of spacing of testing time will individually constitute different experimental designs that may lead to estimators with varying degree of bias (or lack of it) and mean squared error of the TOPE estimator.

2.3 Testing Times and Dose-Time-Response Profiles

Several factors can shape the profile or time trend of effects of acute exposure to potential neurotoxic chemicals. Such factors include the type of chemical agent, the administered dose (and route), the endpoint being measured and the timing of measurements. Timing is an important factor because effects of acute exposure to

neurotoxic compounds usually have specific time profiles, with a certain window of time in which maximum effects can be observed (Zhu et al, 2003b). In the FOB protocol, the US EPA considered also these factors in their recommendation that the 2nd of four testing (the minimum required) be conducted at the estimated TOPE while the remaining three times of testing are fixed. The timing of the 2nd measurement can therefore vary depending mainly on prior knowledge, if any, of the TOPE of a particular chemical-endpoint combination.

2.4 Consideration for Optimal Design Theory

2.4.1 General Principles

According to Tobia (2004) a regression model may be used to investigate the relation between a response variable and a number of explanatory variables. In some cases one is able to choose the values of the explanatory variables, i.e. one can choose in which situations observations can be done. Such choice will determine the quality of the experiment. The theory of experimental design governs the quality of the experiment with respect to its effectiveness of providing relevant information about the model.

Using the notation similar to Tobia (2004), let us consider a model with *n* explanatory variables $x_1,\ldots,x_n$. Under a linear relationship the regression model is given by

$$Y_i = \beta_1 f_1(X_i) + \beta_2 f_2(X_i) + \ldots\ldots + \beta_k f_k(X_i) + \varepsilon_i ,$$

and under a nonlinear relationship by

$$Y_i = f(X_i, \beta) + \varepsilon_i$$

An observation Y_i is the sum of the response function $f(X_i, \beta)$ and error term ε_i , with

$X_i = (x_{i1},\ldots.x_{in})$ as the vector of the explanatory variables, and $\beta = (\beta_1,\ldots\ldots, \beta_k)$ as the vector of unknown parameters. The errors ε_i ($i = 1,\ldots.., N$) are, in the simplest case, assumed to have expectation of zero, constant variance, and to be uncorrelated: $V(\varepsilon_i) = \sigma^2$ and $\varepsilon_i \sim N(0, \sigma^2)$.

Next, we can describe a design as follows. The m points in the experimental region where observations will be done are notated as $X_{1*}, X_{2*}, ..., X_{m*}$, where $X_{i*} = (x_{i1}, x_{i2}, ..., x_{in})$. The number of observations at the point X_{i*} is notated as n_i, so we have

$$\sum_{i=1}^{m} n_i = N$$

with an experiment notated as *Exper*(N) where N indicates how many observations are done in the design, and

$$Exper(N) = (X_{1*}, \ldots., X_{m*}; n_1, \ldots\ldots, n_m; N)$$

Under the linear model, the design matrix is N by k-matrix $\mathbf{X}$ where

$$\mathbf{X} = \begin{bmatrix} f_1(x_{1*}) & \cdots & f_k(x_{1*}) \\ \vdots & & \vdots \\ f_1(x_{1*}) & \cdots & f_k(x_{1*}) \end{bmatrix}_{n_1} \begin{bmatrix} f_1(x_{2*}) & \cdots & f_k(x_{2*}) \\ \vdots & & \vdots \\ f_1(x_{2*}) & \cdots & f_k(x_{2*}) \end{bmatrix}_{n_2} \begin{matrix} \\ \vdots \\ \\ \end{matrix} \begin{bmatrix} f_1(x_{m*}) & \cdots & f_k(x_{m*}) \\ \vdots & & \vdots \\ f_1(x_{m*}) & \cdots & f_k(x_{m*}) \end{bmatrix}_{n_m}$$

For the least squares estimator $\hat{\beta} = (\hat{\beta}_1,\ldots..,\hat{\beta}_k)$ we have

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

where Y is the vector of observations, $Y = (Y_1, \ldots\ldots, Y_N)$, $\boldsymbol{M} = \boldsymbol{X}^T\boldsymbol{X}$ is referred to as the information matrix and **X** is termed the design matrix. If the matrix M is not degenerative, then the matrix $\boldsymbol{M}^{-1}(\mathbf{X}; \beta, \varepsilon)\sigma^2 = Cov(\hat{\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2$, is the dispersion matrix or the variance-covariance-matrix of the best linear estimator of $\hat{\beta}$ (Federov, 1972).

The information matrix depends on the choice of the design **X** and choosing an optimal design means that we have to choose an **X**, say $\mathbf{X}_*$, independent of β and error terms, which makes some real-valued function $\Phi\{\boldsymbol{M}(\mathbf{X})\}$ as large as possible, that is best for all (β, ε). We can say that $\mathbf{X}_*$ is Φ-optimal (Silvey, 1980).

The D-, A-, E- and G-optimality are described briefly as follows.

1. The D-criterion considers the generalized variance, i.e. the determinant of the information-matrix. So a D-optimal design is a design for which the determinant of the information-matrix is made as large as possible.

2. G-optimality is concerned with the variance of a predicted future observation at a given point x_o : $(1 + x_0^T(\mathbf{X}^T\mathbf{X})^{-1}x_0)\ \sigma^2$. The design objective is to minimize the variance.

3. A-optimality considers the trace of the matrix $(\mathbf{X}^T\mathbf{X})^{-1}$. An A-optimal design minimizes the value of $tr(\mathbf{X}^T\mathbf{X})^{-1}$ so that the sum of the marginal variances of the estimators is minimal.

4. E-optimality aims to maximize the eigenvalues of the matrix $(\mathbf{X}^T\mathbf{X})^{-1}$.

2.4.2 Optimal Design under Nonlinear Model

Following the general principles for the linear regression, we now consider our example: Linear-exponential model, a nonlinear case with two predictor variables. The model $y_i = f(t_i, d_i) = A + Bt_i\, d_i\, exp(-K_e t_i) + \varepsilon_i$ is given in section 2.1 and is a special case of the general nonlinear model

$$Y_i = f(X_i, \beta) + \varepsilon_i$$

where $\beta=(A, B, K_e)$ are unknown parameters. Note that the parameter K_e is of special interest because it determines the time of peak effect (TOPE); X_i is a vector of time t_i and dose d_i for the ith observation; and the error terms are independent normal with constant variance: $\varepsilon_i \sim N(0, \sigma^2)$.

The problem of seeking estimates becomes more complicated when the function $f(X_i, \beta)$ is non-linear in β. Using the *Gauss-Newton method*, a Taylor series expansion can be used to approximate the nonlinear regression model with linear terms and then employ ordinary least squares to estimate the parameters (Neter, 1996). Taking a first order Taylor approximation of mean response function $f(X, \beta)$ at the estimate $\hat{\beta}$, we have

$$Y_i - (f(X_i, \hat{\beta}) - \hat{A}\, f_1(X_i, \hat{\beta}) - \hat{B}\, f_2(X_i, \hat{\beta}) - \hat{K}_e\, f_3(X_i, \hat{\beta})$$

$$= A\, f_1(X_i, \hat{\beta}) + B\, f_2(X_i, \hat{\beta}) + K_e\, f_3(X_i, \hat{\beta}) + \varepsilon_i$$

where $f_j = \dfrac{\partial f}{\partial \beta_j}$,

and X is our experimental setting of $t*d$ combination.

From the standpoint of numerical approximation, the design matrix is determined by $f_1(X, \hat{\beta})$, $f_2(X, \hat{\beta})$, *and* $f_3(X, \hat{\beta})$, *with*

$f_1(X, \beta) = 1$; $f_2(X, \beta) = td\ exp(-K_e t)$; and $f_3(X, \beta) = -Bt^2 d\ exp(-K_e t)$.

Here we find that the function $f_2(X, \beta)$ includes K_e while $f_3(X, \beta)$ includes both K_e and B. Unlike in the case of linear regression model, the functions here are dependent on the parameters, and so is the design matrix. The implication is that the optimal design measures are actually dependent upon the true value of β as well as the model. The solution to optimal design is obtained iteratively and necessarily begin with initial or starting values for the regression parameters A, B and K_e .

With the approximate approach, we have a design matrix **D** of partial derivatives now playing the role of the **X** matrix (Neter, 1996). Similarly the form of **D** is

$$\mathbf{D} = \begin{bmatrix} 1 & td\exp(-K_e t)_1 & -Bt^2 d\exp(-K_e t)_1 \\ \vdots & & \vdots \\ 1 & td\exp(-K_e t)_1 & -Bt^2 d\exp(-K_e t)_1 \\ 1 & td\exp(-K_e t)_2 & -Bt^2 d\exp(-K_e t)_2 \\ \vdots & & \vdots \\ 1 & td\exp(-K_e t)_2 & -Bt^2 d\exp(-K_e t)_2 \\ \vdots & \vdots & \vdots \\ 1 & td\exp(-K_e t)_m & -Bt^2 d\exp(-K_e t)_m \\ \vdots & & \vdots \\ 1 & td\exp(-K_e t)_m & -Bt^2 d\exp(-K_e t)_m \end{bmatrix} \begin{matrix} \} n_1 \\ \\ \} n_2 \\ \\ \\ \} n_m \end{matrix}$$

where $n_1 = n_2 = \ldots.. n_m = 50$, $m = 4$; and $1, 2, \ldots m$ correspond to t_1, t_2, t_3, and t_4.

In a typical design setting without constraints, the least squares estimator $\hat{\beta}$ will be given as

$$\hat{\beta} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{Y}$$

where $\hat{\beta}$ is a vector of the least squares estimated regression coefficients. The variance-covariance matrix of $\hat{\beta}$ is

$$\text{Var}(\hat{\beta}) = (\mathbf{D}^T\mathbf{D})^{-1}\,\boldsymbol{\sigma}^2$$

In this case therefore we are looking for designs $\mathbf{D}_*$ which will maximize the optimality function $\Phi\{\boldsymbol{M}(\mathbf{D}), \beta\}$, that is, $\mathbf{D}_*$ may depend on β other than through the information matrix only. Although it could be problematic, the initial or starting values for the parameters would have to be found.

In this thesis, optimal design takes some special constraints. We are interested only in designs with second testing time t_2 to be determined while everything else is fixed. In defining our experimental region, we are constrained by the FOB design protocol so we will focus on finding the optimum timing of 2nd of four repeated measurements with the rest three testing times fixed. Five dose groups with their dose values were pre-decided. Therefore we define our experimental region as follows:

$t = 0, t_2, 24, 168$; where $0 < t_2 =< 20$

d = 0, 0.75, 1.5, 3.0, and 6.0

Also we want to be able to allow for dose group-specific variances.

In summary, the design considered here is a special case in which we maximize C′MC. The C matrix selects particular components of the co-variance matrix. Our focus of interest is to find an optimal design $\mathbf{D}\{f(K_e)\}_*$ for the purpose of estimating TOPE as a function of one of the unknown model parameters K_e while acknowledging that this optimality is also dependent on the parameters β. In addition we would like the optimal

design to accommodate heteroscedasticity. However, instead of seeking algorithms that would enable us to construct the appropriate design measures, we opted for a relatively more empirical approach by doing simulation studies.

2.5 Simulation Rational

In order to evaluate a design, we assume an underlying dose-time-response relationship is given and we generate data according to the dose-time-response relationship and normal random error. The data are simulated under a chosen design. The simulated data are then used to estimate the model parameters under which simulation was done.

A number of experimental designs are possible with respect to the choice of dose and time. However, these designs vary in their capability of revering the information about the true parameters. Thus, we wished to perform a test to determine which of these different designs would produce the best estimate(s) of the population parameters. Our target parameter was the time of peak effects (TOPE). For this purpose, we simulated different designs by generating the FOB data based on the "true" model. An efficient design should allow estimation of the parameters to yield estimates as close to the true value as possible and as reliable as possible. For every simulated design, we fitted the same true model to the data via nonlinear mixed-effects modeling to obtain an estimate of TOPE. The simulation and model fitting process were replicated N times under each design.

The designs which produced the best TOPE estimates were determined based on the bias and mean squared error (MSE) statistics obtained from replicated estimates. The

efficiency of each design was evaluated with specific optimality criteria as follows. The absolute relative bias must be less than 5% of underlying TOPE and/or the design must be associated with the minimum MSE. The minimum MSE was determined both by computation and graphical illustrations. We placed more emphasis on the MSE as a single measure because it combines the effect of bias and sampling variation of the estimator. We applied the concept of coefficient of variation (CV) to relate the MSE to the underlying TOPE. We therefore devised a modified coefficient of variation (mCV) which was computed as a ratio of squared root MSE to the underlying TOPE. This measure was used to compare the variability of different designs based on 2000 replications for each design. Practical designs were determined as those of minimum mCV and/or mCV of no more than 15%.

2.6 Experimental Designs

We employed a number of different experimental designs that essentially were variants of the EPA/IPCS Collaborative Study design (Moser et al, 1997a). These designs differ only with respect to the 2nd testing time point. The design for each acute exposure experiment in the EPA/IPCS Collaborative Study was as follows:

Sample size = 50 rats: 5 dose groups with 10 rats per group

Testing times per rat: t_1 = 0 hr, t_2 = TOPE (hr), t_3 = 24 hr, t_4 = 168 hr

Total number of observations = 200

In line with the EPA/IPCS study protocol above, we fixed the dose levels and also fixed the three testing times t_1, t_3, and t_4 at baseline, 24 and 168 hours post exposure respectively. In order to investigate how the choice of second testing time would affect

the estimation of TOPE, we let t_2 vary between designs. Specifically, a sequence of 30 different designs was chosen with t_2 values ranging between 0.2 hr and 20 hr. Table 2.1 illustrates the EPA/IPCS Collaborative Study design and three of thirty test designs (first, second and thirtieth). It means that the designs allowed for comparison with the one that used the true TOPE as its 2nd testing time.

Table 2.1 Testing Times for EPA/IPCS Collaborative Study Design and Candidate Designs

Testing Times	*EPA/IPCS Study Design*	Design #1	Design #2	…….	Design #30
t_1	*0 hr*	0 hr	0 hr	……	0 hr
t_2	*Estimated TOPE (hr)*	0.2 hr	0.4 hr	……	20 hr
t_3	*24 hr*	24 hr	24 hr	……	24 hr
t_4	*168 hr*	168 hr	168 hr	……	168 hr

Table 2.2 shows all the 30 unique designs represented by the table columns, and of which their t_2 range from 0.2hr to 20hr. These designs were applied to three dose-time-response models coupled with different combinations of functional domain and experiment (or chemical).

Three dose-response models were considered: Linear-Exponential, Toxico-Diffusion (Zhu et al, 2003a) and Rational Function models. Parameter values were taken from previous models fit to the real datasets (Zhu et al. 2003a,b), except for the Rational Function model to which reasonable parameter values were simply assigned. Simulations were based on the following three selected combinations of design, response variable, and model:

1. Acute TET exposure experiment / Linear-Exponential model / Activity domain composite score
2. Acute DDT experiment / Toxico-Diffusion model / Neuromuscular domain composite scores
3. Acute TET exposure experiment / Rational Function model / Activity domain composite scores

Table 2.2 Table of Four Testing Times[1] (Hours) by Thirty Different Designs[2]

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
t_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_2	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2	2.5	3	3.5	4	5
t_3	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24
t_4	168	168	168	168	168	168	168	168	168	168	168	168	168	168	168

	16	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
t_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_2	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
t_3	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24
t_4	168	168	168	168	168	168	168	168	168	168	168	168	168	168	168

1. Testing times designated as t_1, t_2, t_3 and t_4 occur at 0hr, t_2, 24hr and 168hr respectively.
2. The designs numbering from 1 to 30 are designated by their respective t_2 ranging from 0.2hr to 20hr

2.7 Simulation Steps

Our simulation scheme is illustrated below using the first combination: Activity domain scores in conjunction with the Linear-Exponential model. Simulations of the other two combinations were conducted similarly.

2.7.1 Step 1: Define the Dose-Response Model and Population Parameters

Table 2.2 was taken from Zhu et al (2003b) and it shows the results of the Linear-Exponential model fit to the Activity domain composite scores from the acute TET experiment of the EPA/IPCS Collaborative studies (Moser et al, 1997b). We assumed that the fitted model represents a true dose-response relationship in a hypothetical population of rats, that is, the estimated model parameters were taken as the true values for this population.

Based on this "true" model, the TOPE equals 1/ K_o (= 6.16 hrs). The model specification accommodates heteroscedasticity with dose-specific standard error given by 0.2822, 0.2822*1.2884, 0.2822*1.3858, 0.2822*1.2953, 0.2822*2.548 for the five groups of dose=0, 0.75mg, 1.5mg, 3.0mg, and 6.0mg, respectively. The model also contains a random intercept (standard error= 0.1697) for each rat to allow for between-rat variation.

Table 2.3 Acute TET Exposure Study: Activity Scores with Linear Exponential Model[1,2] Fit

Parameter	Value	Std.Error	DF	t-value	p-value
A	1.2738	0.0412	147	30.9129	<.0001
B	0.1755	0.0211	147	8.3261	<.0001
K	0.1623	0.015	147	10.8451	<.0001
Variance Estimate					
Dose	0	0.75	1.5	3	6
StdErr	1	1.2884	1.3858	1.2955	2.5479
Random effects	A	Residual			
Std Dev	0.1697	0.2822			
Model Selection Criteria:	AIC	BIC	logLik		
	248.3158	277.9555	-115.1579		

1. Model=A+B*dose*time*exp(-K_e*time)
2. Distinct variance assumed for each dose group, and the dose-specific standard error is in proportion to that of control

2.7.2 Step 2: Generate Datasets for Each Design

For each design, we simulated an experiment consisting of 10 rats in each of five dose groups. Each rat was tested at 4 time points yielding a total of 200 observations in each experiment. Data were generated based on the following mixed-effects model:

$$y_{ijk} = f(\theta_{ik}, dose_i, time_j) + error_{ijk}\ ,\ i = 1,2,..5;\ j = 1,2,3,4,\ k = 1,2\ldots10$$

In the model θ_{ik} includes also random effects (intercepts) that are additive to the population parameters. The response values were obtained by evaluating the function at the true parameter values, simulated random effects, and the design points of dose and time given in Tables 2.2 & 2.3. The final outcome values were obtained by further adding to the response values the simulated random effects and errors. For the linear exponential model, for example, the outcome is,

$$y_{ijk} = A + a_{ik} + B*dose_i*time_j*exp(-K*time_j) + error_{ijk}$$

Simulations of random effects and errors were accomplished by using computer generated random numbers from specified distributions as follows:

Random effects: This is unique to individual rat. Therefore for 50 rats, 50 random numbers were generated from a normal distribution with zero mean and standard deviation=0.1697 corresponding to that of random intercepts from the fitted model (table 2.2).

Random error and heteroscedasticity: Random error is associated with individual observation and variances are unique to individual dose groups. Therefore, for 40 observations in each of 5 dose groups, 40 random numbers were generated from a normal distribution with a zero mean and standard deviation specific for that dose-group.

Replication: We replicated N=2000 datasets under each design. N was established by allowing it to increase until fitted TOPE was stable. We found N = 2000 satisfactory for all designs in this study.

2.7.3 Step 3: Estimate Parameters

The underlying model was fitted to each simulated dataset to get estimates for parameters A, B and K. TOPE was computed from the estimate of K (TOPE = 1/K). This simulation process resulted in a sample of estimates (2000 replications, convergence rates of about 80% or greater) for each of A, B, K, and the TOPE. Based on this sample estimates, the followings were computed

Bias = sample mean of the TOPE estimates – the "true" TOPE,

Mean Squared Error (MSE) = $Bias^2$ + Sample Variance of the TOPE Estimates,

Relative bias = 100*bias/true TOPE

and

Modified coefficient of variation (mCV) = 100*sqrt(MSE) / true TOPE

2.7.4 Case Large Variance

In order to see the impact of random error on the design, the simulation procedure **(**Steps 1-4) was repeated for the TET/Activity/Linear-Exponential model setting employing larger variation (StdErr=2.0; 78-82% convergence). This illustrative example would enable us to evaluate each study design for this experiment under extreme variability of population dose response profiles.

2.8 Simulation under Two Additional Models

Simulation was conducted for two additional models: Toxico-Diffusion model and Rational Function model. This would allow us to evaluate the sensitivity of designs across model types. The followings are specific information about the models derived from previously fitting the Toxico-Diffusion model to the real data and from simply assigning parameter values to the Rational Function model.

2.8.1 Toxico-Diffusion Model: Acute DDT Experiment / Neuromuscular Domain

Here we had the opportunity to explore a different member of the same family of models as well as a different exposure agent and neurobehavioral domain. Table 2.4 taken from Zhu et al (2003b) shows the toxico-diffusion model fit to the commonality scores of the neuromuscular domain in the acute DDT experiment. The estimated parameter values were used to simulate data under the toxico-diffusion model. From the table, the TOPE estimate directly computed from 1/K is 4.7hr. This value was assumed to be the 'true' TOPE parameter for this setting.

2.8.2 Rational Function Model: Acute TET Experiment / Activity Domain

This model is a simple rational function developed solely for the purpose of this thesis. Unlike the other two models, it was never before fitted to the real FOB data. The reason for inclusion of the model was to further examine the sensitivity of designs to models. In using this model, we were able to assign parameter values such that 1) the model reasonably describes a dose-time-response profile similar to that observed in the original FOB data, and 2) we would have the opportunity to fit a model to simulated

datasets from a population profile with a relatively small 'true' TOPE of 2 hr, the value recorded in the pilot study of the acute TET experiment. (Moser et al, 1997b)

The Rational Function model was specified as follows:

$$\text{model} = A + B*\text{dose}*\text{time}/(K + \text{time}^2), \text{ where TOPE} = \sqrt{K}$$

Two sets of population parameters and sigma were simulated and fitted with Rational Function model as specified in the Table 2.5. The set of parameters (Case 1) describes a response profile similar to that of the exponential model fit to TET/activity scores from the highest dose group. The second set (Case 2) describes still a similar profile but the 'true' TOPE is set lower at 2hr.

Table 2.4 Toxico-diffusion Model[1,2] fit to Neuromuscular Scores of Rats Exposed to DDT[3]

Parameter	**Value**	**Std.Error**	**DF**	**t-value**	**p-value**
A	1.3388	0.0307	147	43.57	<.0001
B	0.0187	0.0073	147	2.58	0.0109
C	0.0108	0.00667	147	1.61	0.1089
K_e	0.2129	0.0327	147	6.52	<.0001
Variance Estimate					
Dose Group	0	10.9	21.8	43.5	87
StdErr	1	0.8859	1.0332	1.6252	1.2037
Random effects A:					
	Intercept	Residual			
StdDev:	0.1042	0.2785			

1. Model=A+B*Dose*Time*exp(-K_e*Time)/ (1+C*Dose*Time*exp(-K_e*Time))
2. Distinct variance assumed for each dose group, and the dose-specific standard error is in proportion to that of control
3. The IPCS/EPA Collaborative Study

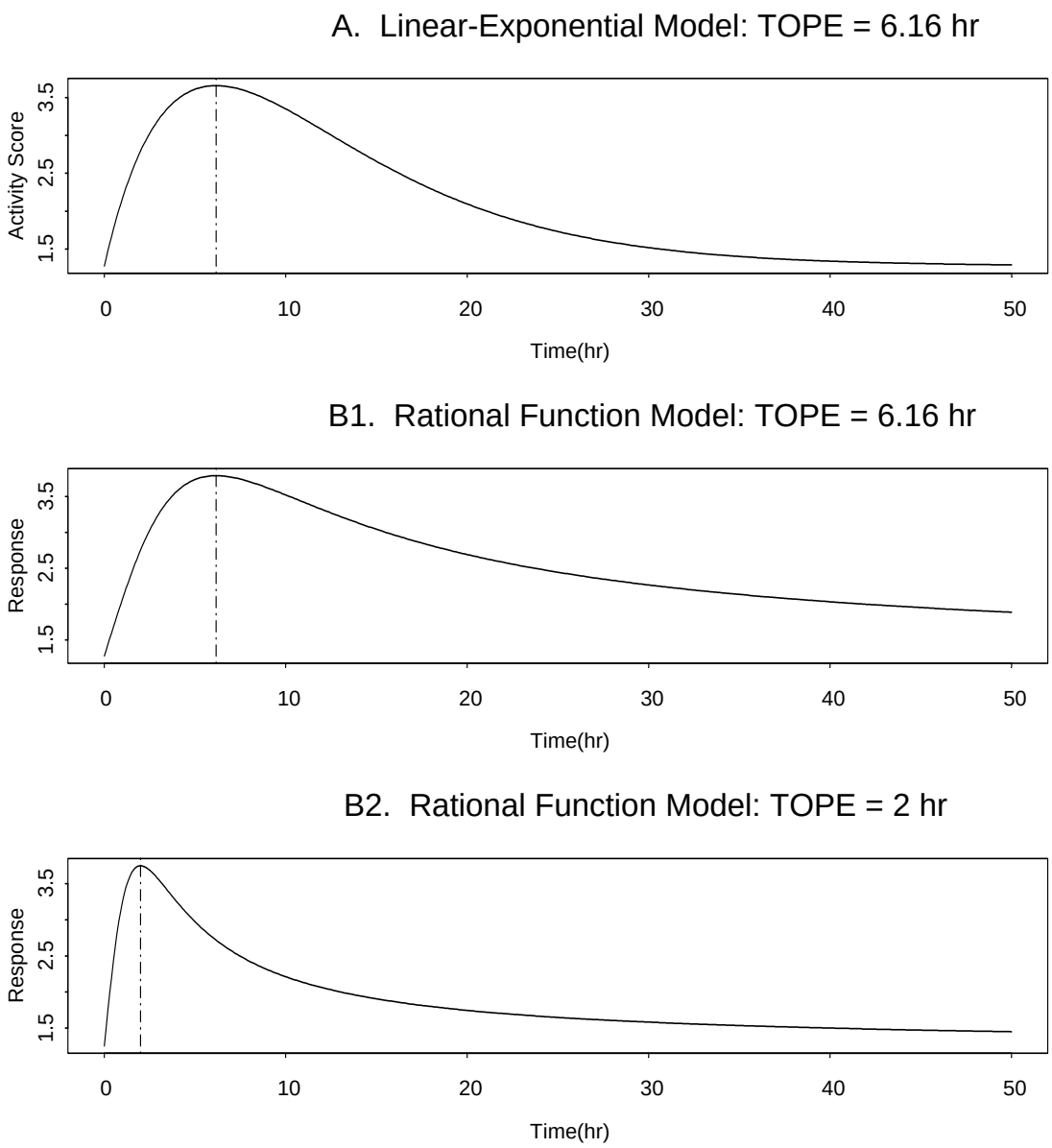
Table 2.5 Rational Function Model Specifications

	Parameter coefficients			Sigma (σ)	TOPE (hr)
	A	B	K		
Case 1	1.27	5.167	37.95	1.0	6.16
Case 2	1.25	1.67	4	0.3	2

In Figure 2.1 the Linear-Exponential (LE) and Rational Function (RF) models are compared by their theoretical dose-response profiles for a fixed dose. In the Figure, plot A displays the theoretical curve for the LE model fit to the Activity scores for the highest dose group in the acute TET experiment (Moser et al, 1997b; Zhu et al, 2003b). Plots B1 & B2 (Figure 2.1) are the dose-response profiles as described by RF model under the assigned parameter values of Case 1 and Case 2, respectively.

In the simulations, the two cases of RF model were used as the base models. The standard deviation was specified as $\sigma = 1.0$ across all dose groups in Case 1 where TOPE=6.12 hr. In Case 2 with TOPE=2.0 hr we specified $\sigma = 0.3$.

Figure 2.1 Theoretical Response-Time Profiles for the Highest Dose[1] Group Based on Comparison Models[2]: Activity Scores of Rats Exposed TET



1. Maximum exposure dose was 6 mg in the acute TET experiment of IPCS/EPA Collaborative Study (Moser et al, 1997b)
2. Linear-Exponential Model = A+B*dose*time*exp(-K_e*time): (A, B, K_e) = (1.27, 0.17, 0.16)
   Rational Function Model = A + B*dose*time/(K + time^2):
   (A, B, K) = (1.27, 5.167, 37.95) for TOPE=6.16 hr
   (A, B, K) = (1.25, 1.67, 4) for TOPE = 2 hr

CHAPTER 3

RESULTS

3.1 Acute TET Experiment: Activity Domain / Linear-Exponential Model

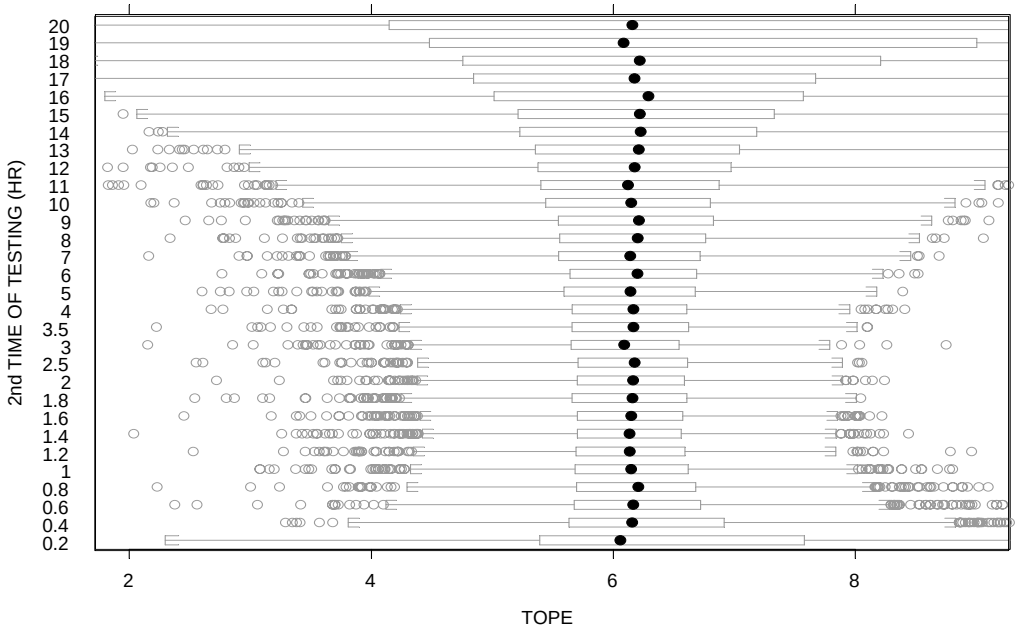
The simulation results are summarized here in this section according to the pattern of standard deviation. Two variance patterns were simulated. In the first category the random effect and dose group specific random errors in the original fitted data were simulated. Specifically, the standard deviations were 0.28, 0.36, 0.39, 0.37, and 0.72 for the five dose groups respectively. In the second category, a large constant value of standard deviation was set at 2.0 for all dose groups.

3.1.1 Distribution of TOPE Estimates

The rate of convergence among 2000 replications was recorded for each of the 30 designs. The convergence rate was greater than 95% in all cases.

Thirty boxplots (one boxplot of 2000 TOPE estimates per design) are displayed graphically side-by-side in Figure 3.1 to show the spread of TOPE estimates for individual designs. The designs with t_2 between about 0.5 hr and 6 hr appeared to have relatively smaller spread for the TOPE estimates. As the second testing time point continues to increase beyond the underlying TOPE=6.16 hr, the TOPE estimator becomes more variable. These findings suggested that the designs which have their 2nd testing times at or before the TOPE are robust to the estimation of TOPE.

Figure 3.1 Boxplots of TOPE Estimates[1] Across Designs[2] for Acute TET Experiment: Simulated Activity Scores[3] with Linear-Exponential Model Fit

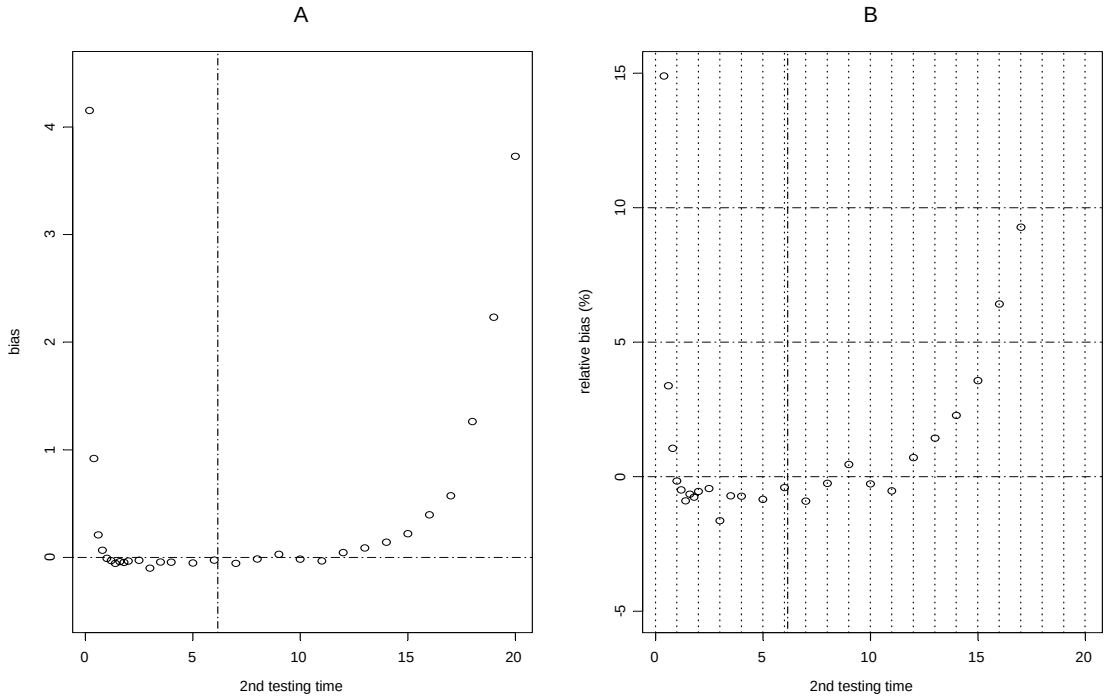


1. One boxplot per design with 2000 replications of TOPE estimates
2. Each of 30 designs is designated by the value of its 2nd time of testing along y-axis
3. Simulated variance pattern is equivalent to that obtained from the original FOB data: different variance per dose group (σ = 0.28, 0.36, 0.39, 0.37, and 0.72), random intercept per subject (σ = 0.17).

The limits of x-axis have been reduced for clarity.

Figure 3.2 shows bias (A) and relative bias (B) of the TOPE estimator. It is seen here that the relative bias of TOPE estimates was less than 5% for the designs with t_2 of 0.6 - 15 hr and was greater than 5% but less than 10% for designs with t_2 between 0.6 and 17 hr. Furthermore, the Figure also shows a likely positive bias associated with t_2 that is either much smaller or larger than the underlying TOPE.

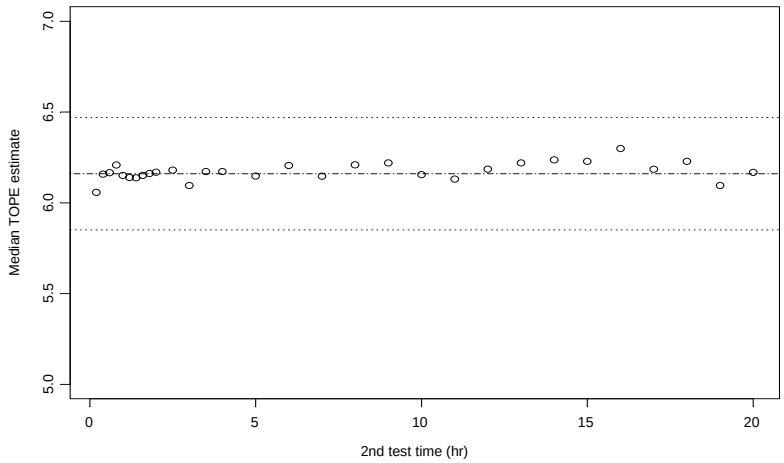
Figure 3.2 Bias and Relative Bias of TOPE Estimates[1] across Designs for Acute TET Experiment: Simulated Activity Scores[2] with Linear-Exponential Model Fit



1. Bias for each design is computed using the mean of 2000 replicates of TOPE estimates. Relative bias = 100*bias/true TOPE.
2. Simulated variance pattern is equivalent to that obtained from the original FOB data: different variance per dose group (σ = 0.28, 0.36, 0.39, 0.37, 0.72), random intercept per subject (σ = 0.17). Each of 30 designs is designated by the value of its 2nd time of testing along the x-axis. Vertical Dash-dot line passes through 2nd testing time at the true TOPE (6.16 hr). Horizontal dashed-lines from bottom mark 0%, 5% and 10% relative bias (B). Vertical axis of B has been reduced to enhance clarity.

From Figure 3.1 we observe that the distributions of TOPE estimates for the designs appear to be normal when t_2 values are in the mid-range but are likely positively skewed when t_2 is smaller or larger. When a distribution is positively skewed, its mean is greater than the median and the median then becomes a more robust measure of the center of the distribution. Therefore the profile of the median TOPE estimates in Figures 3.1 & 3.3 provides a supplementary picture of potential bias of TOPE estimates across designs.

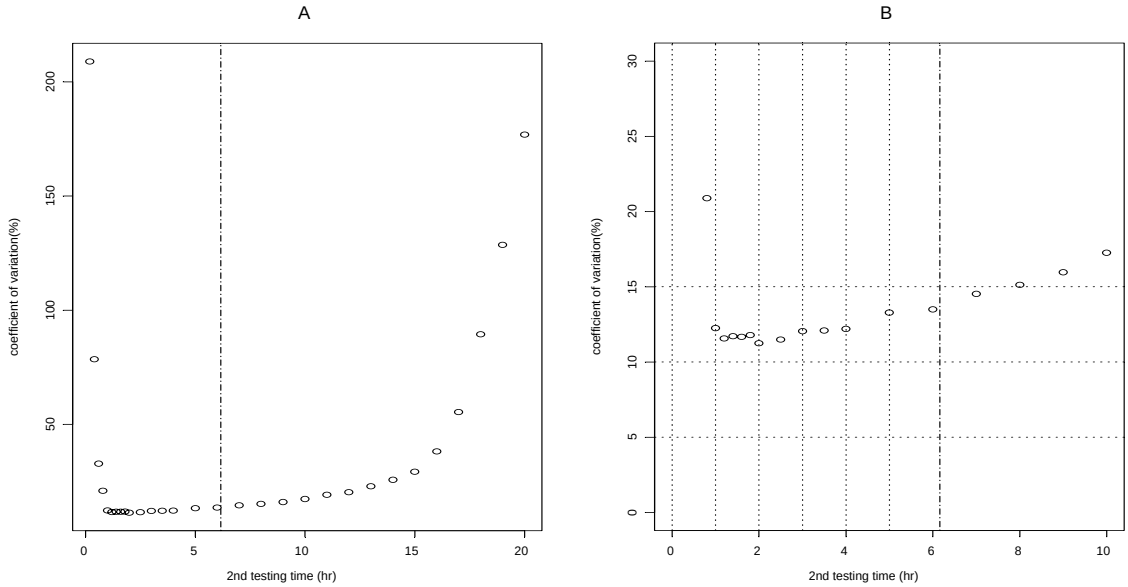
Figure 3.3 Median TOPE Estimates[1] across Designs for Acute TET Experiment: Simulated Activity Scores[2] with Linear-Exponential Model Fit



1. Median of 2000 replicates of TOPE estimates.
2. Simulated variance pattern is equivalent to that obtained from the original FOB data: different variance per dose group (σ = 0.28, 0.36, 0.39, 0.37, 0.72), random intercept per subject (σ = 0.17).
Each of 30 designs is designated by the value of its 2nd time of testing along the x-axis.
Dash-dot line marks the true TOPE (6.16 hr).
Dotted lines mark the upper and lower 5% margins of underlying TOPE.

Coefficient of variation (mCV) is plotted against designs in Figure 3.4. Plot A displays values corresponding to the whole t_2 spectrum (i.e. all designs) while in plot B the focus is on designs with t_2 less than 10 hr. The lowest mCV of 11.3% (corresponding to the lowest MSE of 0.481) was associated with the design of t_2 = 2 hr. However those designs with t_2 between 1 and 7 hours had mCV less than 15% (Figure 3.4B) and these designs were most precise in their estimation of TOPE.

Figure 3.4 Plots of Coefficient of Variation (mCV)[1] Across Designs [2] for Acute TET Experiment: Simulated Activity Scores[3] with Linear-Exponential Model Fit



1. MSE for each design was computed from the mean and the variance of 2000 replicates of TOPE estimates. mCV = 100*sqrt(MSE) / true TOPE
2. Plot A displays all 30 designs. The upper limit of y-axis has been reduced in plot B for clarity.
3. Simulated variance pattern is equivalent to that obtained from the original FOB data: different variance per dose group (σ = 0.28, 0.36, 0.39, 0.37, 0.72), random intercept per subject (σ = 0.17). Vertical Dash-dot line passes through 2nd testing time at the true TOPE (6.16 hr).

3.1.2 Distribution of TOPE Estimates: Large Variance

The case of large variability in dose response was also considered in order to assess the impacts that such large variability might have on designs. We specified a large constant variance (σ = 2.0) across all dose groups in the true model. This was expected to be a challenge to modeling considering that this variation constituted about three times the standard deviation (σ = 0.72) of the most variant dose group in the original dataset. Expectedly, relatively low percent convergences (68-78%) were recorded for the designs associated with both smaller and larger values of the second test times. However designs

with t_2 values of between 0.6 hr and 15 hr recorded convergence fractions of between 78% and 82%.

Table 3.1 Convergence across Designs[1] for Acute TET Experiment: Simulated Activity Scores (Large Variance)[2] with Linear-Exponential Model Fit

Design (t2 (hr))	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2
Convergence (%)	69	76.95	78.45	78.35	76.25	78.45	78.5	79.3	81.05	80.1
Design (t2 (hr))	2.5	3	3.5	4	5	6	7	8	9	10
Convergence (%)	79.1	80.2	80.85	79.6	78.95	80	80.55	81.5	79.05	80.8
Design (t2 (hr))	11	12	13	14	15	16	17	18	19	20
Convergence (%)	80.25	80	78.9	79.7	79.55	76.7	76.65	74.6	71.85	68.9

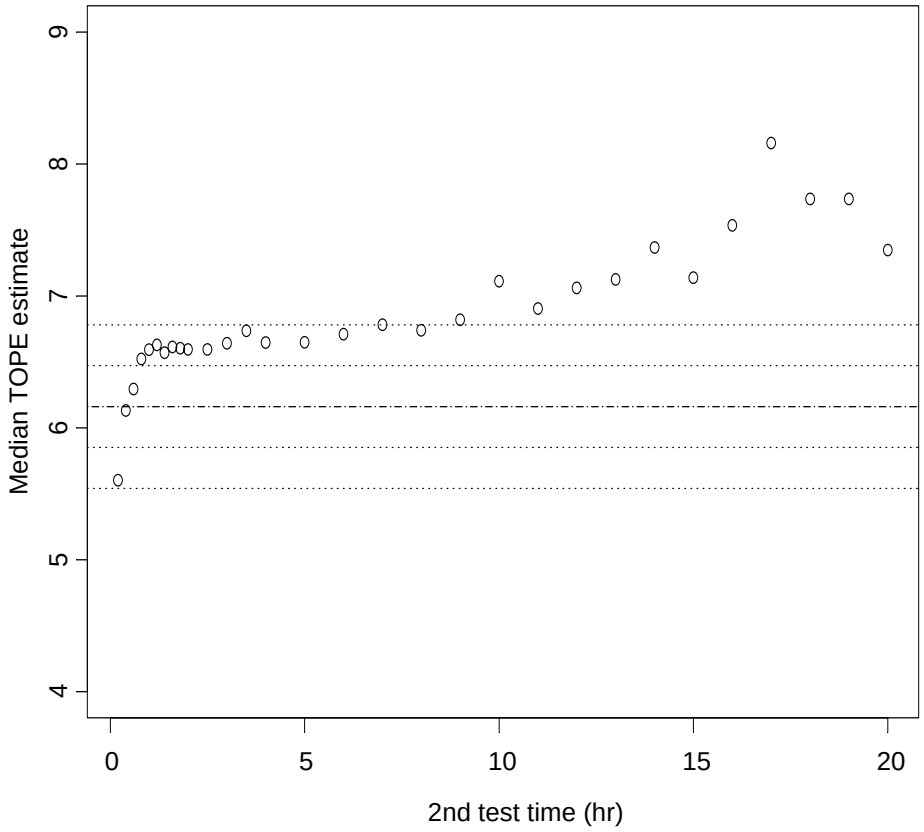
1. Each design is designated by the value of its 2nd time of testing.
2. Constant variance ($\sigma = 2.0$) across dose groups

In Figure 3.5 we see that the median TOPE estimates were within 5% margin of the true TOPE only for designs with t_2 of 0.4 to 0.6 hr. We also see from the figure that the median estimates were within about the 10% margin for designs with t_2 of less than about 10 hr. On the other hand, Figure 3.6 shows that all of the designs were positively biased in their estimation of TOPE; the bias was about 10% above the underlying TOPE when t_2 was between 2 and 5 hr.

Figure 3.7 reveals that the replicated TOPE estimates were generally positively skewed for designs with larger and smaller (to a less extent) values of t_2. Because of the positive skewness, the means were generally greater than the medians and this may offer an explanation for some disparity in profiles across designs seen from Figures 3.5 and 3.6. The bias displayed in Figure 3.6 may be due in part to the skewness of the

distributions and, as a result, the profile of median estimates shown in Figure 3.5 may provide complimentary information.

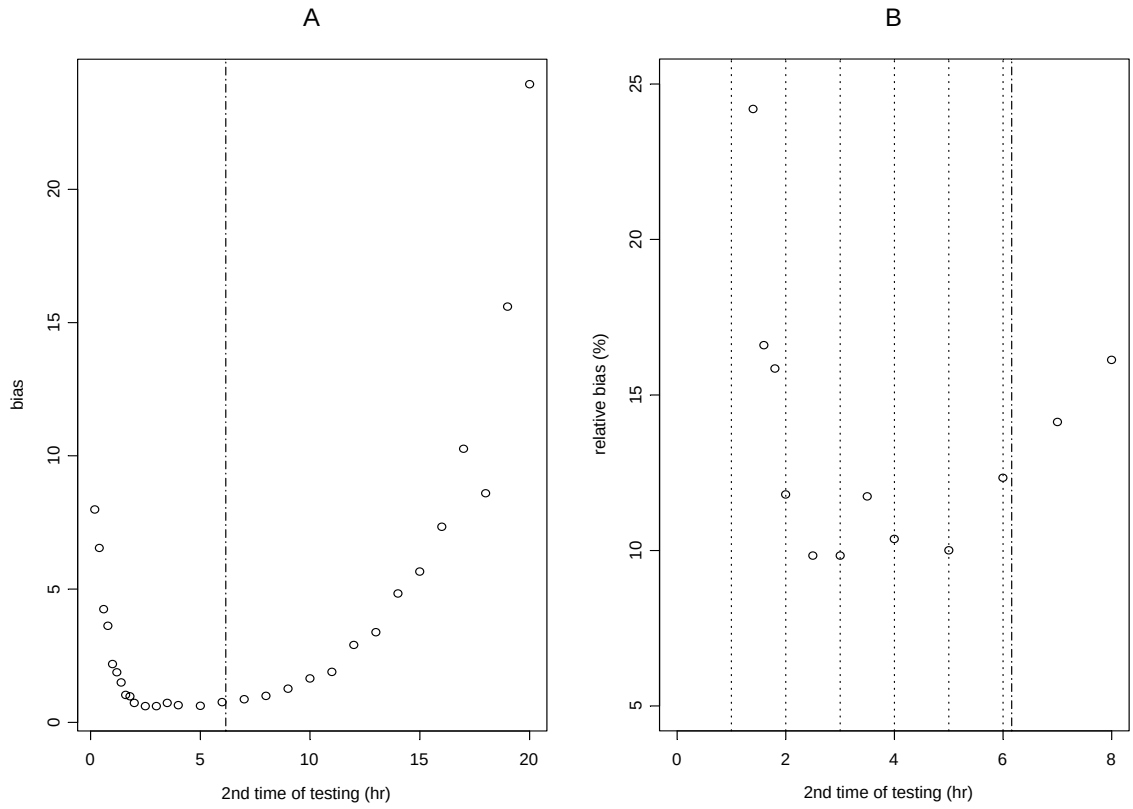
Figure 3.5 Median TOPE Estimates[1] Across Designs[2] for Acute TET Experiment: Simulated Activity Scores (Large Variance)[3] with Linear-Exponential Model Fit



1. Median of 2000 replicates of TOPE estimates.
2. The designs are designated by the values of their 2nd time of testing along the x-axis.
3. Constant variance ($\sigma = 2.0$) across dose groups

Dash-dot line marks the true TOPE (6.16 hr).
Dotted lines (from bottom) mark the lower 10%, 5%, and upper 5% and 10% margin of the underlying TOPE.

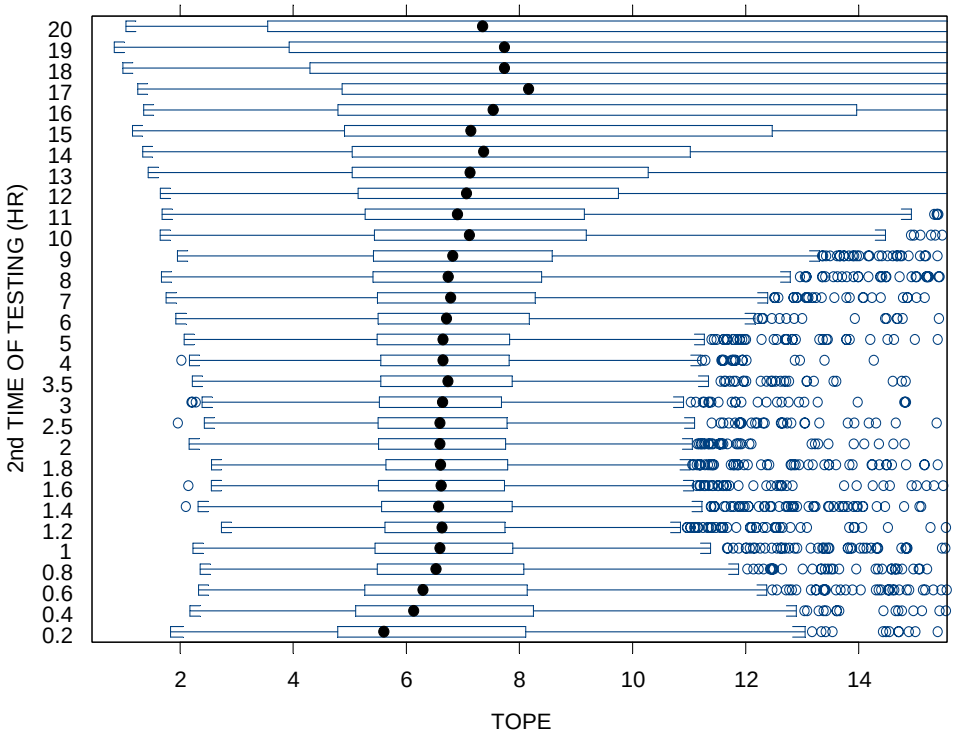
Figure 3.6 Bias & Relative Bias of TOPE Estimates[1] Across Designs[2] for Acute TET Experiment: Simulated Activity Scores (Large Variance)[3] with Linear-Exponential Model Fit



1. Bias for each design was computed using the mean of 2000 replicates of TOPE estimates. Relative bias = 100*bias/true TOPE.
2. The 30 designs (not all is shown on these plots) are designated by the value of their 2nd time of testing along the x-axis.
3. Constant variance ($\sigma = 2.0$) across dose groups

Dash-dot line passes through 2nd testing time at the true TOPE (6.16 hr).
Dotted-lines mark intervals on x-axis.
Limits of both axes in plot B have been reduced to enhance clarity.

Figure 3.7 Boxplots of TOPE Estimates[1] Across Designs[2] for Acute TET Experiment: Simulated Activity Scores[3] (Large Variance) with Linear-Exponential Model Fit



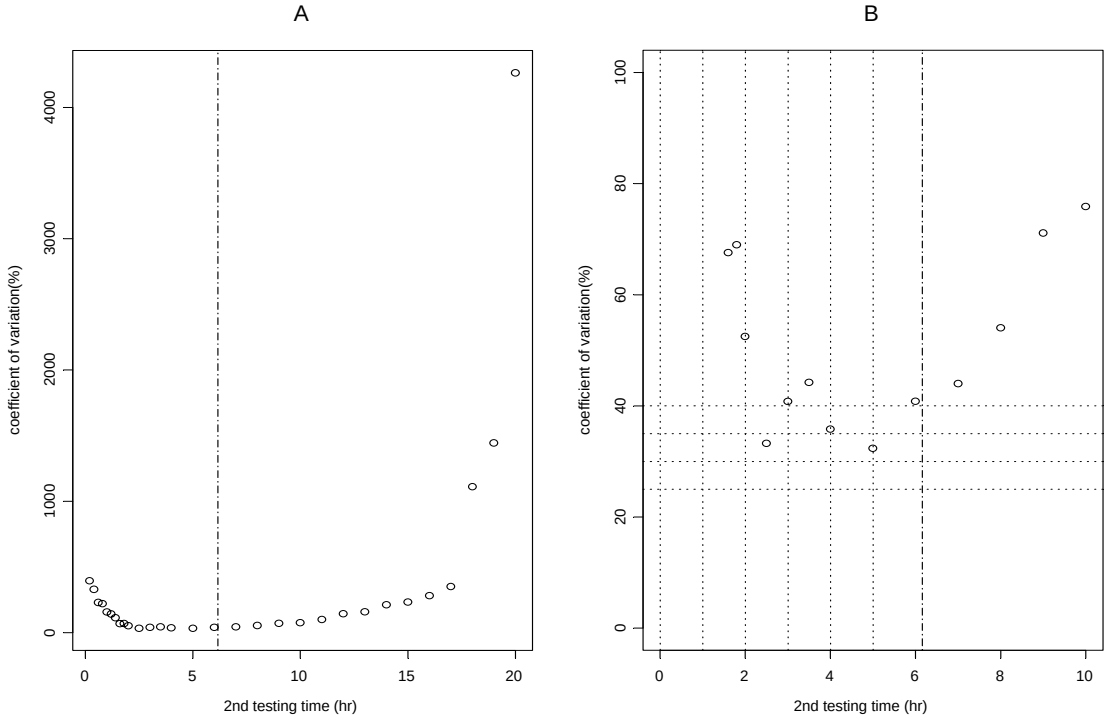
1. One boxplot of 2000 replications of TOPE estimates per design
2. Each of 30 designs is designated by the value of its 2nd time of testing along y-axis
3. Constant variance ($\sigma = 2.0$) across dose groups

Upper limit of x-axis has been reduced for clarity.

It should be noted, however, that on the downside, the estimates from these designs were associated with relatively wide spread. The mCV profile of all designs shown in Figure 3.8A and the expanded form in Figure 3.8B both appear to indicate that designs with t_2 ranging from 2.5 hr to 6 hr are associated with the lowest mCV's with the minimum value of 32.4% occurring at t_2 of 5 hr. Although the mCV trend within this t_2 range (2.5 - 6 hr) appears to be unstable (Figure 3.8B), the smallest spread coupled with

least skewness have been demonstrated for designs in this t_2 range. In contrast, the mCV is about three times as large as the case of small variance (refer to section 3.1.1).

Figure 3.8 Coefficient of Variation (mCV)[1] Across Designs[2] for Acute TET Experiment: Simulated Activity Scores[3] (Large Variance) with Linear-Exponential Model Fit



1. MSE for each design was computed from the mean and the variance of 2000 replicates of TOPE estimates. mCV = 100*sqrt(MSE) / true TOPE
2. Designs with mCV larger than 100% were excluded in this figure for better display of mCV
3. Constant variance ($\sigma = 2.0$) across dose groups compared to most variant dose group ($\sigma = 0.7$) in original data.

Vertical dash-dot line passes through 2nd testing time at the true TOPE (6.16 hr). Dotted lines form grids to aid data point localization.

3.2 Acute DDT experiment: Neuromuscular Domain/ Toxico-Diffusion Model

The DDT experiment is another experiment of the IPCS/EPA Collaborative Study (Moser et al, 1997b). The population parameters, plus random and error variances were obtained from the results (refer to Table 2.4) of a prior fit of the Toxico-Diffusion model

to the dataset (Zhu et al, 2003b). Accordingly, dose group specific variances (σ = 0.28, 0.25, 0.29, 0.45, 0.34) were employed for simulation.

Since the underlying TOPE in the DDT experiment was 4.7 hr, the spacing of t_2 was slightly modified here. A total of 31 designs with t_2 ranging from 0.2hr to 17hr were employed in this phase (Table 3.2). The convergence rate was greater than 96% for all simulated designs when fitting with the Toxico-Diffusion model.

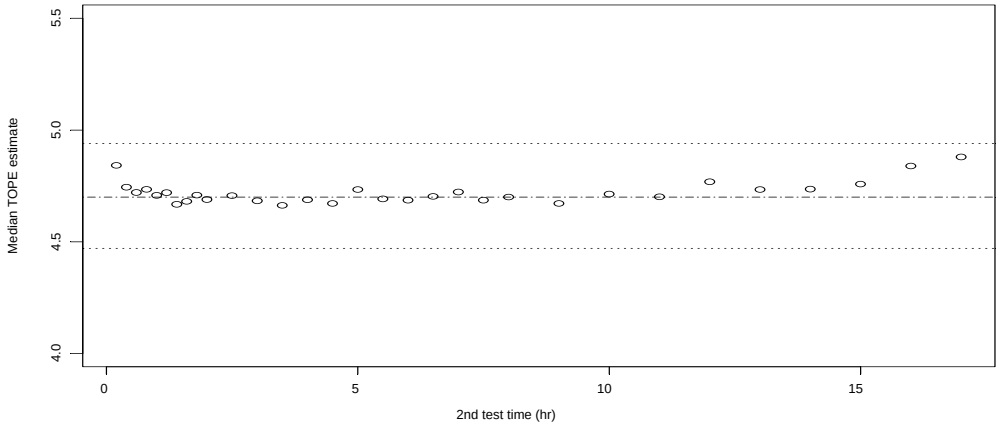
Table 3.2 Table of Designs with Unique 2nd Testing Times for Acute DDT Experiment

design #	1	2	3	4	5	6	7	8	9	10	
2nd test time	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2	
design #	11	12	13	14	15	16	17	18	19	20	
2nd test time	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	
design #	21	22	23	24	25	26	27	28	29	30	31
2nd test time	7.5	8	9	10	11	12	13	14	15	16	17

Figure 3.9 again shows the median TOPE estimates for all designs, which were consistently within 5% margin of the underlying value of 4.7hr. The relative bias (Figure 3.10) was less than 5% for t_2 between 0.4 hr and 12 hr, and less than 10% for t_2 between 0.2 hr and 14 hr; for t_2 = 15 hr and beyond, the bias became increasingly large, reaching more than 30% at t_2 = 17 hr.

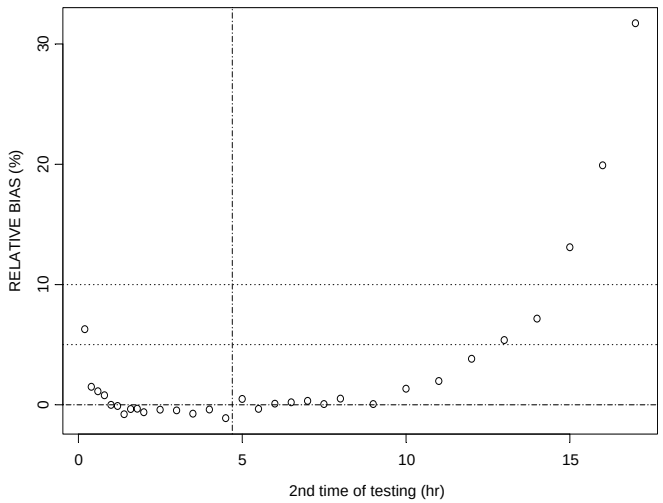
Figure 3.11 shows that the distribution of TOPE estimator here is not much different from the previous boxplots (refer to Figure 3.1). Figure 3.12 clearly demonstrates that the designs with t_2 fixed at 0.8 - 4 hr were associated with mCV of 15% or less.

Figure 3.9 Median TOPE Estimates[1] Across Designs[2] for Acute DDT Experiment: Simulated Neuromuscular Scores[3] with Toxico-Diffusion Model Fit



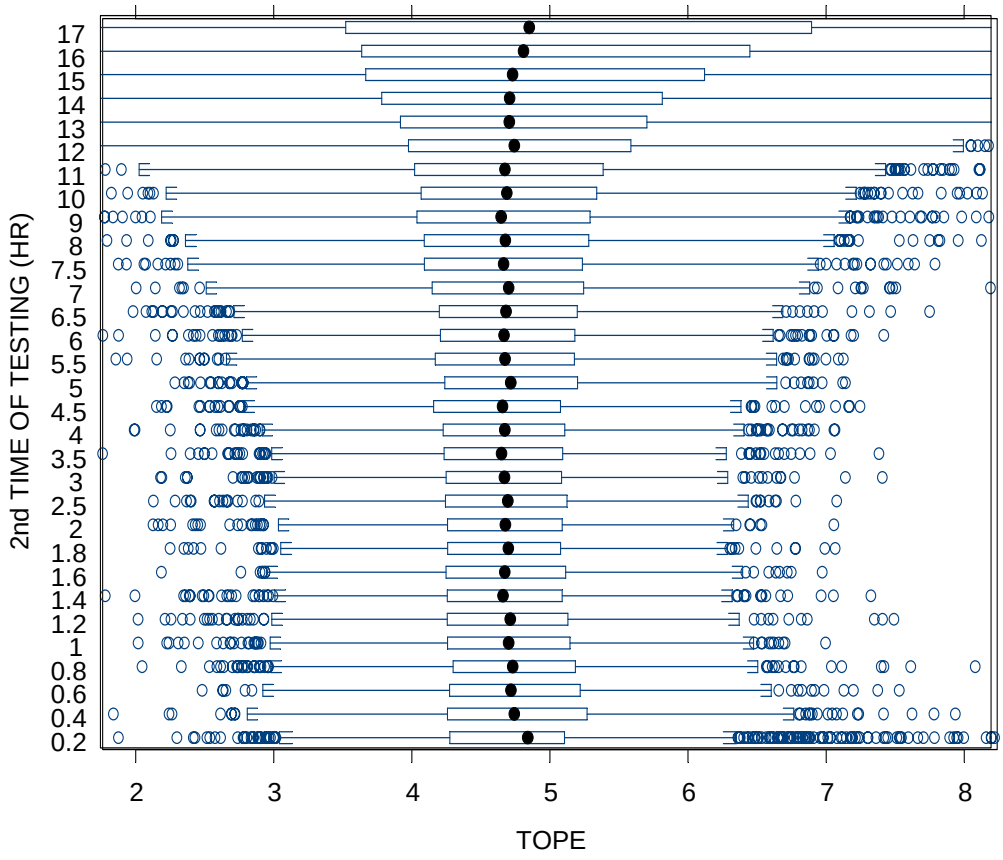
1. Median of 2000 replicates of TOPE estimates.
2. Each design is designated by the value of its 2nd time of testing along the x-axis.
3. Simulated variance pattern is equivalent to that obtained from the original FOB data: different variance per dose group (σ = 0.28, 0.36, 0.39, 0.37, 0.72), random intercept per subject (σ = 0.10). Dash-dot line marks the true TOPE (4.7 hr). Dotted lines (from bottom) mark the lower and upper 5% margins of the true TOPE.

Figure 3.10. Relative Bias of TOPE Estimates[1] Across Designs[2] for Acute DDT Experiment: Simulated Neuromuscular Scores[3] with Toxico-Diffusion Model Fit



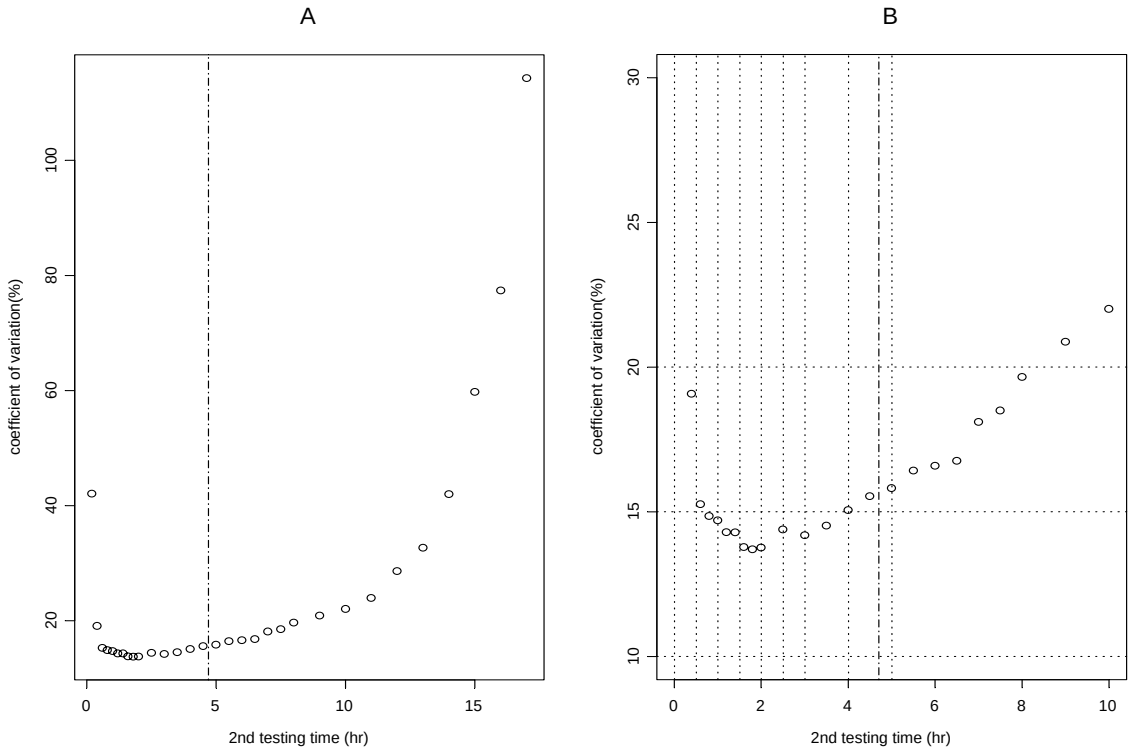
1. Bias for each design was computed using the mean of 2000 replicates of TOPE estimates.
2. The designs are designated by the value of their 2nd time of testing along the x-axis.
3. Simulated variance pattern is equivalent to that obtained from the original FOB data: different variance per dose group (σ = 0.28, 0.36, 0.39, 0.37, 0.72), random intercept per subject (σ = 0.10). Dash-dot line marks the true TOPE (4.7 hr).
Dotted lines (from bottom) mark the 0%, 5% and 10% margins of the true TOPE.

Figure 3.11 Boxplots of TOPE Estimates[1] Across Designs[2] for Acute DDT Experiment: Simulated Neuromuscular Scores[3] with Toxico-Diffusion Model Fit



1. One boxplot per design with 2000 replications of TOPE estimates
2. Each of 31 designs is designated by the value of its 2nd time of testing along y-axis
3. Simulated variance pattern is equivalent to that obtained from the original FOB data: different variance per dose group (σ = 0.28, 0.36, 0.39, 0.37, 0.72), random intercept per subject (σ = 0.10). Upper limit of x-axis has been reduced to aid visualization.

Figure 3.12 Plots of Coefficient of Variation (mCV)[1] Across Designs[2] for Acute DDT Experiment: Simulated Neuromuscular Scores[3] with Toxico-Diffusion Model Fit



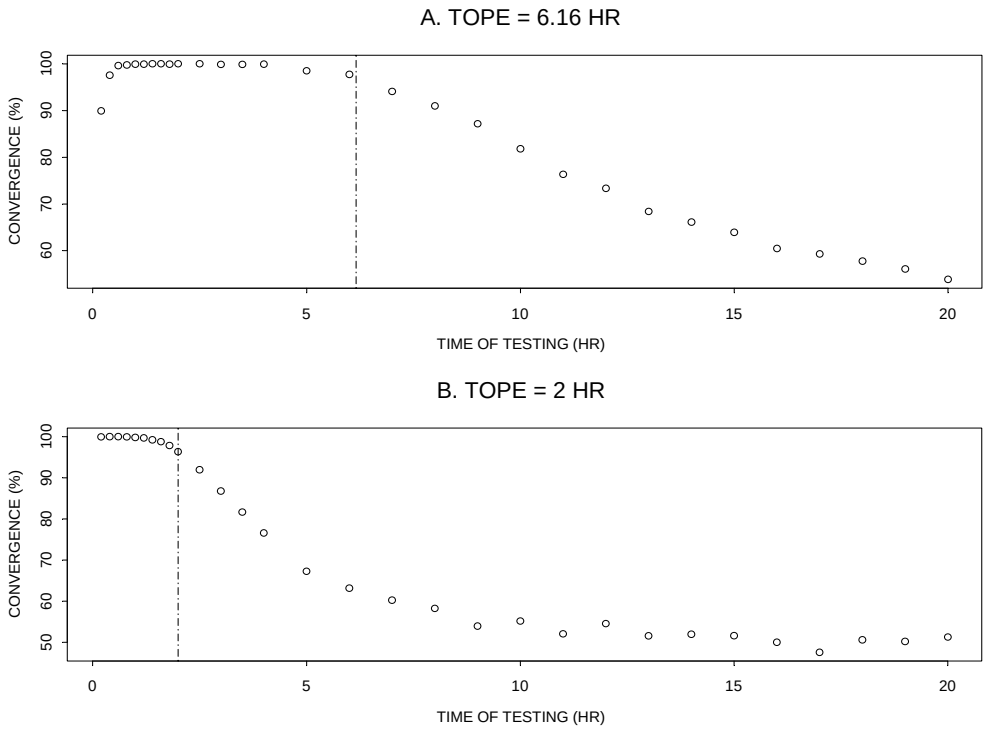
1. MSE for each design was computed from the mean and the variance of 2000 replicates of TOPE estimates. mCV = 100*sqrt(MSE) / true TOPE.
2. Only plot A displays all the 31 designs. The upper limits of both axes have been reduced in plot B for better display of mCV.
3. Simulated variance pattern is equivalent to that obtained from the original FOB data: different variance per dose group (σ = 0.28, 0.36, 0.39, 0.37, 0.72), random intercept per subject (σ = 0.10). Vertical dash-dot line passes through 2nd testing time at the true TOPE (4.7 hr). Dotted lines form grids to aid localization of data points.

3.3 Acute TET Experiment/ Activity Domain: Rational Function Model

This model is based on rational function that may also be used to describe a dose-response profile similar to those demonstrated by the Activity domain in the acute TET experiment. The main purpose of using this model is to test sensitivity of design with respect to models, particularly with an underlying TOPE as small as 2 hr.

Two sets of population parameters and sigma were simulated. For the first case, TOPE = 6.16 hr and standard deviation was specified as $\sigma = 1.0$ across all dose groups. The standard deviation was about 40% larger than that of the most variant dose group ($\sigma = 0.28*2.55 = 0.71$) from the original dataset (Table 2.2). The second case had TOPE = 2 hr with $\sigma = 0.3$ for every dose group. In each of both simulations, the convergence rate of fitting simulated data was greater than 80% under designs of t_2 values less than 10 hr (Figure 3.13A, case 1) and 3.5 hr (Figure 3.13B, case 2).

Figure 3.13 Convergence across Designs[1]: Simulated Activity Scores[2] of the Acute TET Experiment With Rational Function Model Fit



1. Each of 30 designs is designated by the value of its 2nd time of testing along the x-axis.
2. Sigma equals 1.0 for A and 0.3 for B.

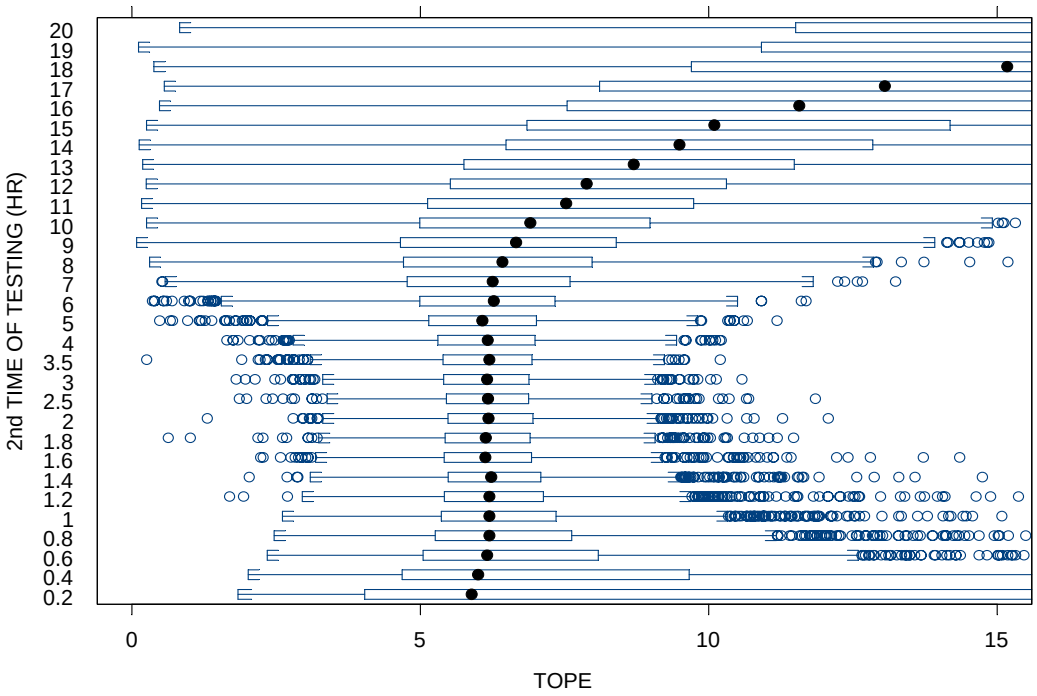
Vertical dash-dot line is the underlying TOPE.

Since in each instance the t_2 range associated with good convergence stretches beyond the underlying TOPE reasonably well in both directions, there is sufficiently wide time window within which to reliably test candidate designs. Therefore non-convergence is not a problem.

3.3.1 Rational Function Model: Case One

Figure 3.14 shows that the spread of the replicated TOPE estimates is relatively

Figure 3.14 Boxplots of TOPE Estimates[1] to Compare Designs[2] for Acute TET Experiment: Simulated Activity Scores[3] (TOPE = 6.16 hr) With Rational Function Model Fit

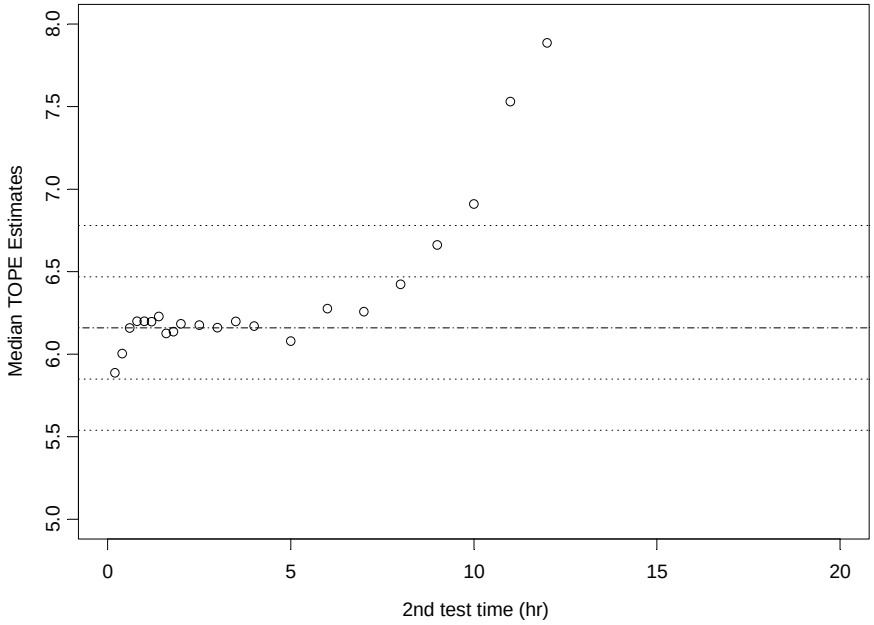


1. One boxplot per design with 2000 replications of TOPE estimates
2. Each of 30 designs is designated by the value of its 2nd time of testing along y-axis
3. Simulated total variance is 1.0 (control group variance for the original FOB data = 0.6)

Upper limit of x-axis has been reduced for clarity.

small, apparently between 1 hr and 5 hr. The median increasingly shifted to the right (increased) starting from t_2 greater than about 9 hr but shifted to the left (decreased) when t_2 was less than about 0.6 hr. In addition, for both of these extreme t_2 values, the skewness increased and the variability of the TOPE estimator became increasingly large.

Figure 3.15 Median TOPE Estimates[1] by Designs for Acute TET Experiment: Simulated Activity Scores[2] (TOPE=6.16 hr) with Rational Function Model Fit



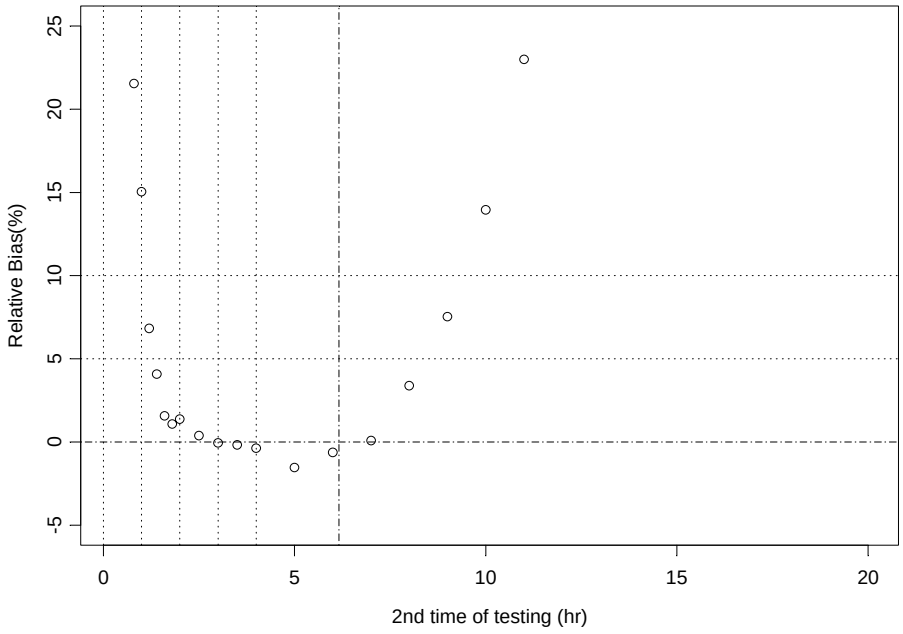
1. Median of 2000 replicates of TOPE estimates.
2. Simulated constant variance per dose group (σ= 1.0).

Each of 30 designs is designated by the value of its 2nd time of testing along the x-axis.
Dash-dot line marks the true TOPE (6.16 hr).
Dotted lines mark the upper and lower 5% & 10% margins of the true TOPE.
Upper limit of y-axis has been reduced for clarity.

In Figure 3.15, the median TOPE estimates for designs within the t_2 ranges of 0.2-8 hr and 0.2- 9 hr are shown to be within the 5% and 10% margins of the underlying TOPE respectively. For the bias, Figure 3.16 shows that the relative bias was no more

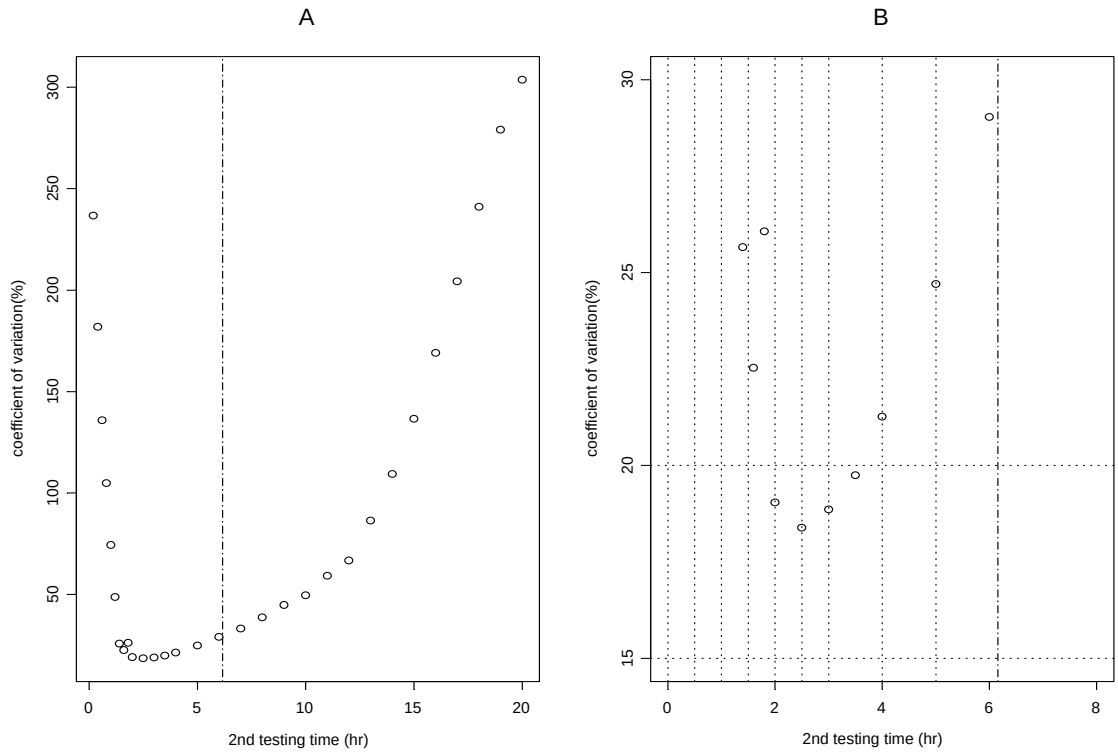
than 5% and 10% for t_2 in the ranges of 1.4 - 8 hr and 1.2 - 9 hr respectively. The mCV for all designs tested were greater than 15% while a lowest mCV of 18.4% was recorded for t_2 of 2.5 hr (Figure 3.17).

Figure 3.16 Relative Bias of TOPE Estimates[1] Across Designs[2] for Acute TET Experiment: Simulated Activity Scores (TOPE=6.16 hr) with Rational Function Model Fit



1. Bias for each design was computed using the mean of 2000 replicates of TOPE estimates.
2. Thirty designs are designated by the value of their 2nd time of testing along the x-axis. Vertical dash-dot line passes through 2nd testing time at the true TOPE (6.16 hr) while dotted lines form grids to aid data point localization.

Figure 3.17 Plots of Coefficient of Variation (mCV)[1] Across 30 Designs [2] of Acute TET Experiment: Simulated Activity Scores[3] with TOPE=6.16 hr and Rational Function Model Fit



1. MSE for each design was computed from the mean and the variance of 2000 replicates of TOPE estimates. mCV = 100*sqrt(MSE) / true TOPE.
2. Designs are designated by the value of their 2nd time of testing along the x-axis. All 30 designs are displayed in A while in B the upper limits of the x- and y-axes have been reduced for clarity.
3. Simulated constant variance per dose group (σ= 1.0).

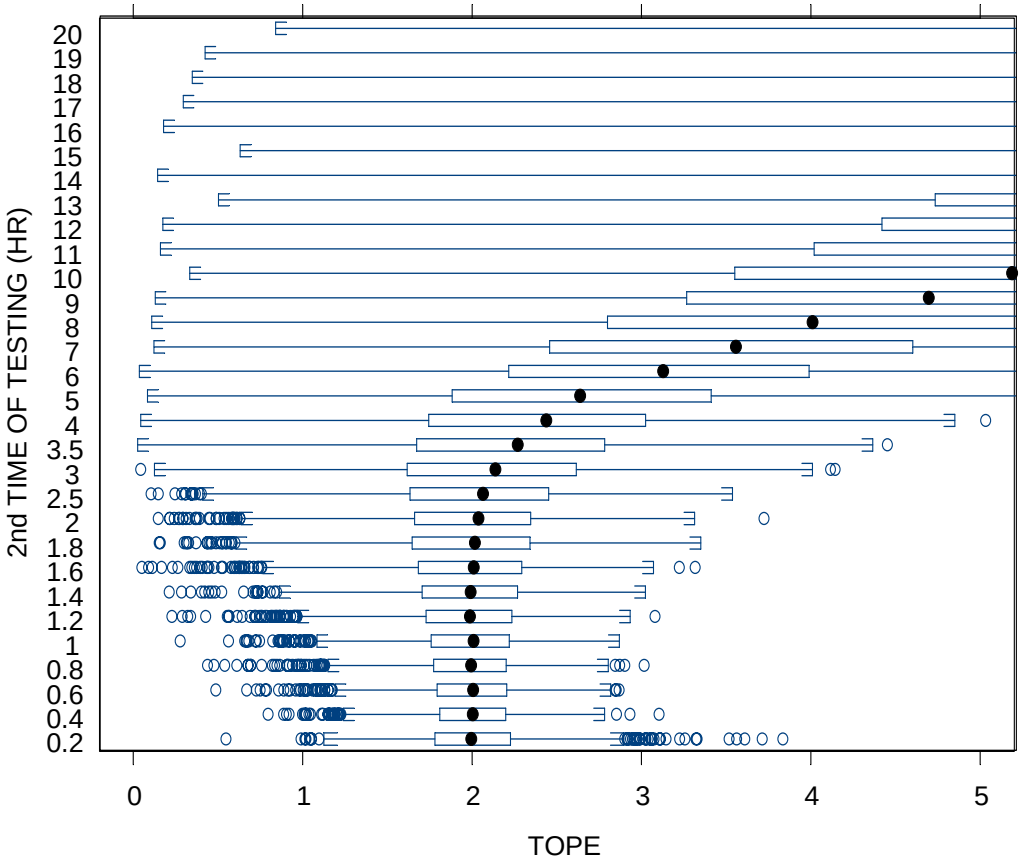
Vertical dash-dot line passes through 2nd testing time at the true TOPE (6.16 hr). Dotted lines form grids to aid data point localization.

3.3.2 Rational Function Model: Case Two

Inspection of Figures 3.18 and 3.19A reveals a progressive shift of the median estimate away from the underlying TOPE as t_2 increased from 2 hr. As t_2 increased, the spread and skewness of the replicated TOPE estimates also increased. Quantitatively, the median TOPE estimates were within the 5% and 10% margins of the true value when t_2

was no greater than 2.5 hr and 3 hr respectively (Figure 3.19 A). However those designs of which t_2 values were 0.2 - 3 hr produced estimates with no more than 5% relative bias. (Figure 3.19 B).

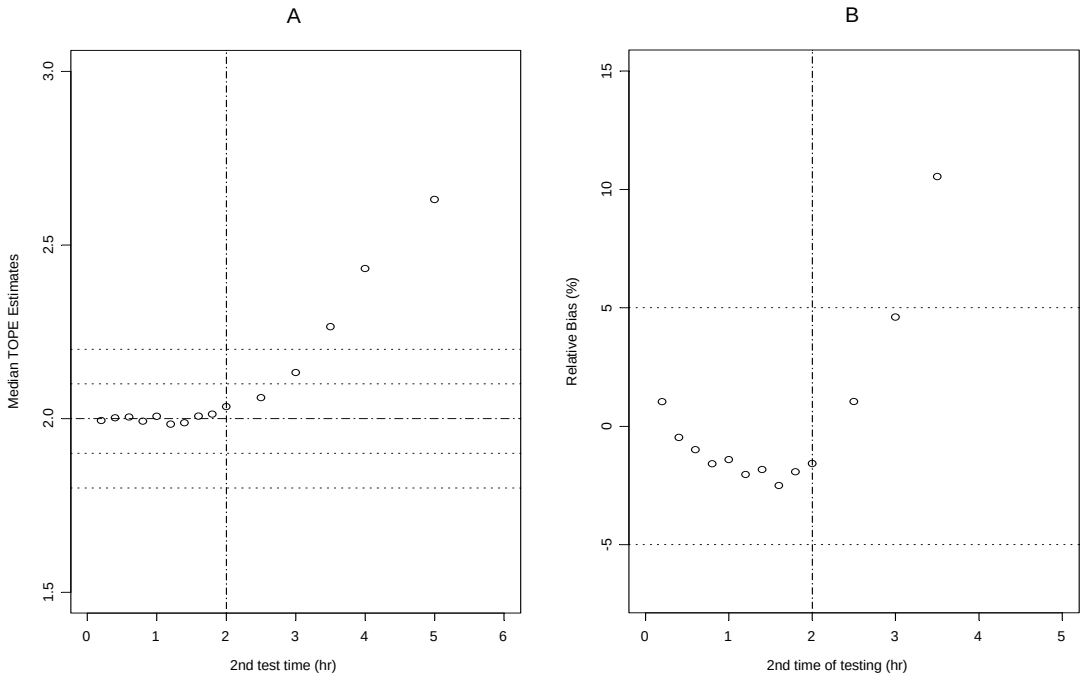
Figure 3.18 Boxplots of TOPE Estimates[1] to Compare Designs[2] for Acute TET Experiment: Simulated Activity Scores[3] (TOPE = 2 hr) With Rational Function Model Fit



1. One boxplot per design with 2000 replications of TOPE estimates
2. Each of 30 designs is designated by the value of its 2nd time of testing along y-axis
3. Simulated constant variance per dose group (σ= 0.3).

Upper limit of x-axis has been reduced for clarity.

Figure 3.19 Median and Relative Bias of TOPE Estimates[1] Across Designs[2] for Acute TET Experiment: Simulated Activity Scores (TOPE=2 hr) with Rational Function Model Fit

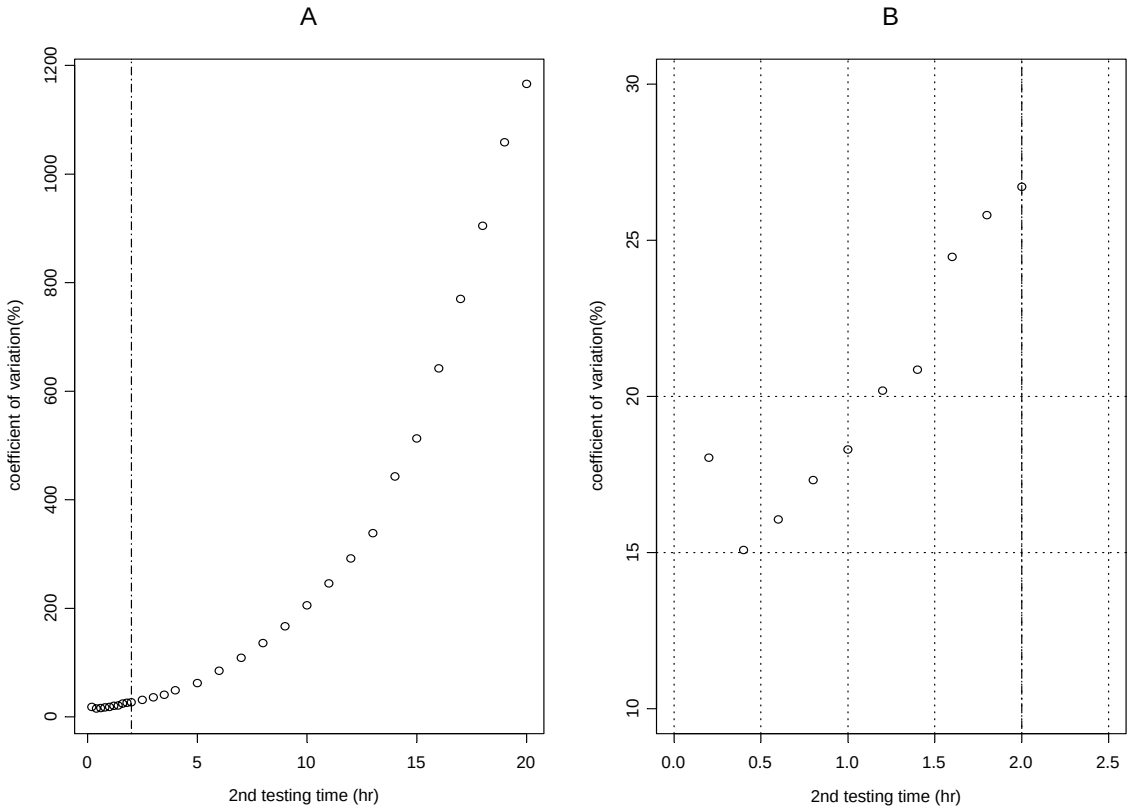


1. 2000 replicates of TOPE estimates.
2. Each of 30 designs is designated by the value of its 2nd time of testing along the x-axis.

Dash-dot line marks the true TOPE (2 hr). Dotted lines mark the upper and lower 5% , 10% (A) and 5% (B) margins of the true TOPE.
The limits of both axes have been adjusted for better display.

All of the designs tested in this case were associated with relatively high mCV (Figure 3.20). The minimum mCV was 15.1% and was recorded for the design t_2 of 0.4 hr (Figure 3.20B).

Figure 3.20 Plots of Coefficient of Variation (mCV)[1] Across 30 Designs [2] of Acute TET Experiment: Simulated Activity Scores[3] (TOPE = 2 hr) with Rational Function Model Fit



1. MSE for each design was computed from the mean and the variance of 2000 replicates of TOPE estimates. mCV = 100*sqrt(MSE) / true TOPE.
2. Designs are designated by their 2nd testing time along the x-axis. All designs are displayed in A while in B the limits of the x- and y-axes have been adjusted to better display mCV.
3. Simulated constant variance per dose group (σ= 0.3).

Vertical dash-dot line passes through 2nd testing time at the true TOPE (2 hr). Dotted lines form grids to aid data point localization.

3.4 Summary of Results and Interpretations

The results are summarized in Table 3.3 below. Each row (category) of the table represents one distinct setting of an experiment with respect to FOB domain and dose response model structure. Thirty (31 for category C) different designs (of different t_2) were evaluated within each setting. In the four categories where each involved 30 designs, the t_2 values tested ranged from 0.2 hr to 20 hr while for category C the range was 0.2 hr to 17 hr. The time window included the underlying TOPE (6.16 hr, 4.7 hr and 2.0 hr) in every instance.

In cases A & C where the simulated dataset share the same variance pattern with the original dataset, designs within a wide range of t_2 (0.6-12 hr) yielded TOPE estimates with no more than 5% relative bias away from the true TOPEs. Similarly, in case E where the underlying TOPE was relatively small (2 hr), only the designs in the t_2 range of 0.2 - 3 hr were able to produce estimates lying within 5% relative bias.

As the variance of the simulated data increased over that of the original dataset, a decreasing number of designs qualify as efficient. For example, in cases A and C where variance is comparable to that in the original dataset, the range of t_2 required to estimate the TOPE to within 10% of bias was in each case 0.6 -17 hr and 0.2 – 14 hr respectively. However in case B (with about 200% increase in standard deviation over the most variant group in case A), a narrower range of t_2 (2-5 hr) was required to achieve the same level of accuracy for estimating the TOPE. Case D is intermediate between A and B with respect to variance (about 40% increase) and qualifying designs (1.2 – 9 hr).

Table 3.3 Summary of Designs for the Estimation of TOPE

Experiment / Domain /Model	Variance Pattern in Simulated Data	True TOPE (hr)	Best Designs Designated by Best 2nd Time of Testing or t_2 (hr)					
			Conver-gence (%)	Median within 5% margin of TOPE	Relative Bias =<5%	Relative Bias =<10%	mCV =<15%	Lowest mCV
A: TET/ Activity/ LE	Different variance per dose group/ Random intercept	6.16	0.2 -20 All designs (>95%)	0.2 -20 All designs	0.6 - 15	0.6 -17	1 - 7	2.0 (11.3%)
B: TET/ Activity/ LE	Large constant variance $\sigma = 2.0$ (~ 200% larger)	6.16	0.6 - 15 (78 - 82%)	0.4 -0.6	none	2 – 5	none	5.0 (32.4%)
C: DDT/ Neuro-muscular/ TD	Different variance per dose group/ Random intercept	4.7	0.2 -17 All designs (>95%)	0.2 -17 All designs	0.4 -12	0.2 - 14	0.8 - 4	1.8 (13.7%)
D: TET/ Activity/ RF	Constant variance $\sigma = 1.0$ (~ 40% larger)	6.16	0.2 -10 (>80%)	0.2 -8	1.4 - 8	1.2 – 9	none	2.5 (18.4%)
E: TET/ Activity/ RF	Constant variance per dose group $\sigma = 1.0$	2.0	0.2 - 3.5 (>80%)	0.2 -2.5	0.2 - 3	0.2 – 3.	0.4	0.4 (15.1%)

Legend:
LE Linear-exponential model
TD Toxico-diffusion model
RF Rational function model
TOPE Time of peak effects
mCV Modified coefficient of variation

Generally the above findings suggest that for the cases considered, in order to produce reasonably accurate estimates of the TOPE to say within 5% of the true value, a design must choose a 2nd testing time not far away from the underlying TOPE. Furthermore they suggest that the presence of wide variability in the data may reduce the capability of designs for accurate estimation of the TOPE and further restricts the choice of t_2 for effective designs to values less than the underlying TOPE. Although the models are different in most of the cases considered here, it is reasonable to expect that variation in data may influence the designs as suggested by our findings.

With respect to mCV, the bias and variance of estimation are combined. There is a direct linear relationship between mCV and MSE with lower values of either indicating high precision of estimation for a given design. Compared to relative bias (=<5%), here a much narrower range of t_2 was consistently required to achieve a desirable level of precision of estimates (mCV=< 15%) irrespective of the model or pattern of variance in the data. Generally where TOPE=6.16 hr, t_2 should be about 2.5 hr in order to attain the smallest mCV which varied between 10% and 30% for the designs considered. For TOPE = 4.7 hr, the minimum mCV of 13.7% was achieved at t_2=1.8 hr, however t_2 would be in the range of about 1-4 hr in order to have mCV of no more than 15%. Similarly, a minimum mCV of 15% was obtained by only one design of t_2=0.4 hr when the underlying TOPE was 2 hr.

Overall, the number of designs with the greatest precision (smallest mCV) is a subset of designs with the highest validity (least bias) in the estimation of TOPE. In the cases considered in the this thesis, the most precise TOPE estimates were produced generally when the second testing time was situated about midway between time zero and

the underlying TOPE. An exception where the most precise design had its second testing time (5 hr) relatively closer to the underlying TOPE (6.16 hr) was category B where variability in the sample was large (σ=2). Here the smallest attainable relative bias (about 10%) and mCV (about 32%) were comparatively larger than those of other categories. It should be recalled that the trends of both measures (relative bias and mCV) across designs were rather unstable in the minimum regions (refer to Figures 3.6 & 3.8). The implication is that for this category, there is probability that the appropriate t_2 for the most effective design could be anywhere from 2 hr to 6 hr, which still leans more to the lower side of the underlying TOPE of 6.16 hr.

In all, the t_2 value of each of these identified effective designs remained smaller than the underlying TOPE. It follows that under various combinations of conditions such as exposure agent, neurobehavioral domain, statistical model, or value of the underlying TOPE, all of the qualifying effective designs seemed to share a common robust feature that the 2nd testing time should be chosen at a point a little earlier than the underlying TOPE in order to achieve robust estimation of the TOPE.

CHAPTER 4

DISCUSSION

The IPCS/EPA Collaborative Study protocol (Moser et al, 1997b) under which the existing FOB data were generated proposed that the 2nd testing in a particular experiment be performed at the time of peak effects (TOPE) for that chemical. The TOPE is derived using two endpoints through a pilot experiment. Since true TOPE may vary with the testing chemical, the dosing level, and the endpoint, the choice of the 2^{nd} testing time can be important in determining the quality of the experiment. This thesis set out principally to find effective designs with respect to the choice of the second testing time point. Through simulation of a set of designs uniquely defined within a range of the 2nd testing time, the most effective designs were selected based on specified criteria. The results of the study showed that many designs are robust against a misspecification of the TOPE choice, and can produce TOPE estimates within a relative bias of 5% margin. These designs are also robust with respect to the criterion mean squared error (MSE) or modified coefficient of variation (mCV); however the range of t_2 becomes narrower because of the inclusion of variance in these criteria. Further, empirical evidences show that these designs prefer to have the 2^{nd} testing time point before the true TOPE. However, it is not clear in general how earlier the second testing time point can be. Further investigation will be helpful before our results can be generalized to a broader situation.

The dose-response models considered in our study dictates that the TOPE is a function of the model parameters, and does not vary with dose level. Our simulation utilizes parameter values derived from several real datasets. For each design, we simulated 2000 replication experiments, and fit the underlying dose-response model to them. The convergence rate was generally high when fitting the dose-response models to simulated datasets. Bias in estimating TOPE was generally negligible for most designs.

Although, based on our findings, there seems to be reasonable latitude allowable around the TOPE for the choice of 2nd testing time in order for the statistical estimate of TOPE to be associated with no more than 5% relative bias, this may not in itself be sufficient or be readily achievable in practice. As our study further shows, those designs with second testing performed at the TOPE may be associated with relatively high MSE or mCV. That means that in a single experiment, there is a high probability for such design to yield a TOPE estimate with more than 5% deviation from the true value. Alternatively, designs in which the second testing were performed at about halfway below the true value of TOPE were credited with the least MSE in our study and therefore can be expected to have the highest probability of producing TOPE estimates within the 5% margins of the true value in a single experiment.

The main interpretation of our findings may be exemplified as follows. Let us consider for example an ideal situation under the proposed IPCS/EPA protocol. Prior to a certain acute exposure experiment, the TOPE was accurately determined to be 4 hr post exposure in a range finding pilot study. As implied by the findings of this thesis, if we conduct the 2nd testing of the experiment at 2 hr (or between 1 and 3 hr) post exposure, the subsequent statistical estimate of the TOPE has a higher probability of being close to

the true TOPE value of 4 hr than if we had conducted the testing at 4 hr. Therefore the timing of the 2nd testing has an impact on the overall capability for statistical estimation of the true TOPE on the basis of a fitted dose-time-response model.

The significance of our findings can be further illustrated by exploring a scenario closer to real under the IPCS/EPA protocol. Major sources of uncertainty in the TOPE estimate obtained from a pilot study include systematic errors or bias (inaccuracy) and random error or statistical variation (imprecision). Therefore, if a pilot study came up with an estimated TOPE = 4 hr, given the uncertainty of estimation we can reasonably assume that the true value could be anywhere between 3 and 5 hr. If the conclusions from our present findings were to apply, then the 2nd testing would be performed at halfway below the estimated TOPE, which would be at 2 hr. In effect the 2nd testing time (2 hr) would be about halfway below the true TOPE (which lies anywhere between 3 and 5 hr). That means this design would be close to optimal in spite of the uncertainty in the TOPE estimate from the pilot study. On the other hand, under the IPCS/EPA protocol, the 2nd testing would have to be performed at 4 hr. Such design might be fairly close to optimal if the true TOPE was between 4 and 5 hr but the design would definitely be even further away from optimal if the true TOPE lied between 3 and 4 hr.

The last scenario above is very conceivable given the fact that the pilot experiment based TOPE estimates obtained for just two FOB measures may not be truly representative of all the 30 FOB response measures both within and across neurobehavioral domains. Hence it is reasonable to expect that the TOPE estimate may be fraught with substantial uncertainty as depicted above. Designing the experiment proper so that the 2nd testing time is about half way below the pilot experiment based

TOPE estimate is therefore recommendable. This may increase the probability of getting TOPE estimate that is close to the true value thereby facilitating effective statistical estimation of the TOPE.

The findings and recommendations of this thesis may have a limited direct application to the OPPTS guidelines released in 1998 (US EPA, 1998b), where the proposed minimum times of testing are before exposure, at TOPE, and 7 and 14 days post exposure. The present study was evaluated under the IPCS/EPA protocol that produced the FOB data, and where times of testing were before exposure, at TOPE, and 1 and 7 days post exposure. If the dosing effects are transient such that the toxic effects are largely washed out between day 1 and day 7, then data collected on day 14 provides very little additional information beyond those data collected on previous testing times. In that situation our findings may be inapplicable to such data generated under the 1998 guidelines. It should be noted though that the inter-individual variability with respect to dose-time-response characteristic that is distributed in a given population of rats should be inherent to that population regardless of under which protocol (whether 1997 or 1998 protocol) observations are made. So far as the statistical models referred to in this thesis adequately fit the dose-time-response trend (e.g. single peak, maximum or minimum; no premature washout) in a given sample, our present findings may be applicable under both protocols. Nevertheless, the potential impact that the difference between data generated under both protocols may have on statistical modeling should be a subject of future study.

In further research it will be useful to investigate whether at least two testing time points surrounding the TOPE may be needed, one earlier and one later. It will also be helpful to assess the impact of TOPE estimate on the variability of Benchmark dose

(BMD) which is related to the TOPE estimate. Such research may help to further quantify the relative contributions of the comparison designs tested in this thesis to the variation in TOPE and BMD estimations.

REFERENCES

Anger, W. K. (1986). Workplace exposures. In Neurobehavioral Toxicology (Z. A. Annau, ed., pp. 331-347. Johns Hopkins University Press.

Crump, K. S. (1984). A new method for determining allowable daily intakes. *Fundamental and Applied Toxicology* **4**, 854-871.

Fedorov, V. (1972). *Theory of optimal experiments*. Academic Press, New York.

Lammers, J. H. C. M., and Kulig, B. M. (1997). Multivariate time of peak effects assessment for use in selecting time of testing in acute neurotoxicity studies. *Neurotoxicology* **18**, 1079-84.

MacPhail, R. C., Tilson, H. A., Moser, V. C., and al, e. (1997). The IPCS Collaborative Study on Neurobehaviroal Screening I. Background and Genesis. *Neurotoxicology* **18**, 925-928.

McDaniel, K. L., and Moser, V. C. (1993). Utility of a neurobehavioral screening battery for differentiating the effects of two pyrethroids, permethrin and cypermethrin. *Neurotoxicology and Teratology* **15**, 71-83.

Moser, V. C., Becking, G. C., Cuomo, V., , and al, e. (1997). The IPCS collaborative study on neurobehavioral screening methods V: results of chemical testing. *Neurotoxicology* **18**, 969-1056.

Moser, V. C., Tilson, H. A., MacPhail, R. C., Becking, G. C., Cuomo, V., Frantik, E., Kulig, B. M., and Winneke, G. (1997). The IPCS collaborative study on

neurobehavioral screening methods II: protocol design and testing procedures. *Neurotoxicology* **18**, 929-938.

Neter, J., Kutner, M., Nachtsheim, C., and Wasserman, W. (1996). *Applied linear statistical model*. WCB/McGraw-Hill, Boston, MA.

Silvey, S. (1980). *Optimal design*. Chapman and Hall, London.

Tobias RD (Retrieved 2004). The structure of optimal design algorithms. http://support.sas.com/rnd/app/papers/optex.pdf.

U.S. EPA (1991). Pesticide assessment guidelines, subdivision F. Hazard evaluation: human and domestic animals. OPPTS Addendum 10. PB91-154617. NTIS, Springfield VA, Washington, DC.

U.S. EPA (1995). Reportable quantity adjustments; final rule, pp. 60(112):30926-30962. Federal Register.

U.S. EPA (1998). Neurotoxicity Screening Battery. OPPTS 870.6200. In Health Effects Test Guidelines EPA 712-C-98-238.

US EPA (1990). Neurotoxicity: identifying and controlling poisons of the nervous system, Vol. OTA-BA-436, pp. 105-144. Office Of Technology Assessment (OTA), U.S. Congress, Washington, DC: Government Printing Office.

US EPA (1998). Guidelines for Neurotoxicity Risk Assessment, pp. 63:26926-26954. Federal Register.

Wackerly, D., Mendenhall, I. W., and Scheaffer, R. (1996). *Mathematical statistics with applications*. Wadsworth Publishing Company, California.

Zhu, Y. (2001). Neurobehavioral toxicity risk assessment: dose-time-response modeling and benchmark dose estimation. Reported to National Center for Environmental Assessment, US EPA.

Zhu, Y. (2003). Mixed-effects toxico-kinetic and toxico-dynamic dose-response models for neurotoxicity risk assessment. *Environmentrics. Under review*.

Zhu, Y., Toyinbo, P., (2003). Analyses of Neurobehavioral Data: Dose-time-response Modeling of Composite Scores. *Unpublished manuscript*.

Zhu, Y., Jia, Z., Wang, W., Gift, J., Moser, V., and Pierre-Louis, B. (2004c). Analyses of the Neurobehavioral Screening Data III: Benchmark Dose Estimation. *Under EPA's peer review*.

Zhu, Y., Toyinbo, P., Woodruff, S., Liu, T., and Moser, V. (2004b). Analyses of Neurobehavioral Data II: Dose-time-response Modeling of Composite Scores. *Under EPA's peers review*.

Zhu, Y., Wessel, M., Woodruff, S., Liu, T., and Moser, V. (2004a). Analyses of Neurobehavioral Screening Data I: Dose-Time-Response Modeling of Continuous Outcomes. *Under EPA's peers review*.