7-18-2005

Dose Time Response Modeling of Neurobehavioral Screening Data: Application of Physiologically Relevant Parameters to Allow for Dose Dependent Time of Peak Effects

Michael Raymond Wessel
University of South Florida

Dose Time Response Modeling of Neurobehavioral Screening Data: Application of Physiologically Relevant Parameters to Allow for Dose Dependent Time of Peak Effects

by

Michael Raymond Wessel

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

Major Professor: Yiliang Zhu, Ph.D.
Getachew Dagne, Ph.D.
Jeff Gift, Ph.D.

Date of Approval
July 18, 2005

Keywords: dose-response, biphasic, neurotoxicity, risk, toxicology

DEDICATION

To Joyce Ann

Who allowed me to chase the snakes, and make my own mistakes,

supporting me all the while. Thanks for the dance.

Michael

Acknowledgments

I am grateful to Dr. Yiliang Zhu, my advisor and major professor, who has given up countless hours of well deserved rest to provide me with his guidance, expertise and the support necessary to complete this thesis.

I wish to express my sincere gratitude to Dr. Jeff Gift of the USEPA for agreeing to serve on my committee and provide invaluable insights regarding this work. I also thank Dr. Gift's colleagues; Dr. Allan Marcus, Dr. Virginia Moser and Dr. Rob Dewoskin for reviewing the content of the thesis and providing their insight on the potential application of this effort to their very important work

I further thank Dr. Getachew Dagne for his constructive comments and willingness to share his insights and joyful personality.

In running out of room I finally thank my family; Mom, Maury, Mark, Craig, Linda, Alex, J.T., Jane, Mathew, Bob, Laura and, unendingly, my wife Megan who bore the brunt of my frustration and still encouraged me to fight on.

With Love,

Mike

Table of Contents

List of Tables

List of Figures

Dose Time Response Modeling of Neurobehavioral Screening Data: Application of Physiologically Relevant Parameters to Allow for Dose Dependent Time of Peak Effects

Michael Raymond Wessel

ABSTRACT

In collaboration with the United States Environmental Protection Agency (USEPA), the University of South Florida Health Risk Methodology Group has developed dose-time-response models to characterize neurobehavioral response to chemical exposure. The application of dose-time-response models to neurobehavioral screening tests on laboratory animals allows for benchmark dose estimation to establish exposure limits in environmental risk assessment. This thesis has advanced dose-time-response modeling by generalizing a published toxico diffusion model to allow for dose dependent time of peak effects. To accomplish this, a biphasic model was developed which adopted the effect compartment model paradigm used in pharmacokinetics/pharmacodynamics to estimate a distributional rate constant to account for dose related variation in the time of peak effect. The biphasic model was able to describe dose-dependent time of peak effects as observed in the data on acute exposure to parathion and adequately predicted the observed response. However, the experimental design appeared insufficient in statistical power to confirm statistical significance for each parameter of interest. Motivated by the question of what design requirement might be necessary to validate the biphasic model, Monte Carlo simulation was adopted. Simulations were performed to assess the efficacy and efficiency of various experimental designs for detecting and evaluating some critical characteristics of the biphasic model, including the TOPE. The results of simulation suggest that the location of measurement times around the TOPE have important implications for assessing the statistical significance of the parameter that describes dose-dependent TOPE and that the mean squared error of the parameter estimator was improved most when testing times were

chosen to bracket the TOPE. While dose dependent time of peak effects has underlying physiological mechanisms such as synergistic or capacity limited kinetics, the biphasic model estimates these physiological properties through a mathematical function which may be physiologically relevant but does not necessarily define physiological mechanisms underlying the response. However, if verified through further testing, the biphasic model may contribute to the USEPA's aim of developing physiologically relevant dose-response models for assessing risk of neurotoxicity with repeated measurements of response.

Chapter One

Introduction

Chemical exposure has become a certainty of human life in the 21st century. In fact, it is difficult to imagine a day goes by in which one is not exposed to a chemical, natural or synthetic, about which there is some uncertainty of risk. Formal attempts to characterize the risks associated with chemical exposure date back to Hippocrates (ancient 400 BC) who developed toxicological principles related to clinical observations on the bioavailability and absorption of common therapies and poisons. Paracelsus (~1500 AD) is credited with the idea that all substances are poisons and it is the dose that determines its potential risk and benefit. In modern times, exponential growth in the production of chemicals occurred as a consequence of the industrial revolution and World War II and in the United States led to the establishment of regulatory agencies such as the Food and Drug Administration and later the Environmental Protection Agency (US EPA). Because of the many natural and synthetic chemicals introduced in today's environment, governmental agencies throughout the industrialized world have become keenly interested in assessing the potential risks to humans from toxic agents (US EPA, 1998). In the

US, the EPA has registered more than 65,000 chemical substances manufactured, imported, or processed in the United States under the Toxic Substance Control Act (TSCA) (US EPA, 1996). Chemical exposure has become one of the ten leading causes of workplace disorder (Anger, 1984) and the potential for adverse effects on the nervous system is becoming common in the workplace as approximately 70 chemicals of known neurotoxic potential have potential exposure to more than 1 million workers (Anger,1990).

1.1 Health Risk Assessment

Environmental risk assessment is an emerging field that relies on three basic assessment principles: exposure assessment, hazard characterization, and risk quantification (McCarty and Mackay, 1993). The EPA's National Center for Environmental Assessment (NCEA) conducts risk assessment for an array of health effects that may result from exposure to environmental agents. This process includes a thorough evaluation of all the available data as well as conducting scientific experiments to understand the relationship between exposure and risk. Historically, these analyses have been done very differently for cancer and non-cancer health effects because of perceived differences in the mechanistic underpinnings of cancer and other toxic effects. As our understanding of the underlying biology of toxic effects has grown, however, the apparent differences between

cancer and non-cancer effects have lessened to the point where it seems reasonable to develop quantitative methods based on similar considerations for all types of health effects, and to make approaches to risk assessment as consistent across health endpoints as our current mechanistic understanding allows (US EPA, 2000). Neurotoxicity risk assessment is one area in particular where the EPA has expressed the need for consistent guidance on how to evaluate data on neurotoxic substances and assess their potential to cause transient or persistent and direct and indirect effects on human health.

1.2 Neurotoxicity

Neurotoxicity is defined as an adverse change in the structure or function of the central and/or peripheral nervous system following exposure to a chemical, physical or biological agent (Tilson, 1990). The central nervous system is particularly vulnerable to chemical insult and has limited ability to regenerate. Functional neurotoxic effects include adverse changes in somatic/ autonomic, sensory, motor and/or cognitive function (US EPA, 1998). The effects can be transient (the organism returns to pre-exposure condition) or persistent (the organism is permanently and adversely changed by exposure). However, even transient effects can signify underlying resultant damage to the organism (US EPA, 1998). Animal studies make up the largest portion of controlled exposure assessments and allow for the

use of high concentrations of chemicals to be administered to achieve responses that may define the mechanism of action as well as the magnitude of response to chemical exposure. Neurotoxic screening tests such as the Functional Observational Battery (FOB) test (Moser et al., 1995, 1997a), along with neuro-physiological, biochemical, neuro-pathological and neuro-endocrinological studies are now being used by the EPA as an overall strategy to detect the full range of chemical induced alterations in the structure and function of the nervous system.

The advancement of neurotoxicity testing methods and experimental design has coincided with advances in statistical methodologies and computer applications to allow for more effective methods of performing risk assessments on neurotoxins. A significant advance in neurotoxicity risk assessment is the use of benchmark dose (BMD) methodologies for establishing safety levels of chemical exposure (US EPA, 2000). While traditional analysis of FOB data has used Analysis of Variance (ANOVA) to set a No Observed Adverse Effects Level (NOAEL), dose-time-response models provide continuous estimation of response over the time course of the study providing beneficial information for BMD estimation.

The University of South Florida's Health Risk Assessment Methodology Group (HRAMG) has been working to develop new

statistical methods for explicit dose-time-response models (DTR) of the FOB data, which may serve as the foundation for benchmark dose estimation. The development of explicit DTR models to describe neurotoxic potential of chemical exposure has progressed from strictly mathematical models such as polynomial models to those incorporating simple toxicokinetics. The potential for physiologic interpretation enhances comparability with other available data and increases confidence in the interspecies extrapolation of results of these screening tests to characterize potential risk to human health. Zhu (2005a) and Zhu et al. (2005b,c) have developed a family of dose-response models and illustrated their application through several published datasets generated from the EPA Superfund study (Moser et al, 1995) and a study conducted in collaboration with the International Program on Chemical Safety (IPCS) (Moser et al. 1997a). These models incorporate basic toxicokinetic principles into a family of mathematical models in consideration of the physical properties underlying responses to chemical exposure observed in the FOB data. Zhu (2005 b,c) found that the toxico diffusion model often satisfactorily described the observed dose-response relationship in FOB data and is useful in application of benchmark dose estimation methods. While the toxico diffusion model has proven robust in describing FOB data, it is limited in describing the full possible range of

dose related response to acute exposure. Most importantly this model is limited by imposing a dose independent Time of Peak Effect (TOPE) across every exposure level.

1.3 Thesis Objectives

The first objective of this thesis was to expand on a published toxico-diffusion function used to model neurobehavioral screening data by considering a situation where the TOPE may take on dose-dependent characteristics. Through the incorporation of a dose-dependent distributional parameter, this thesis proposes a "biphasic diffusion" model and illustrates the utility of this model on the analysis of a FOB dataset on the motor activity of laboratory rats exposed to the organophosphate pesticide parathion.

A second objective of the thesis was to investigate the study design requirements necessary to recover the key characteristics of the biphasic model, especially dose-dependent time of peak effect. Monte Carlo simulation was adopted to explore the potential FOB testing times as well as sample size with respect to the efficiency and effectiveness of recovering key components of the biphasic model. Various designs were compared using statistical measures of bias, power, mean squared error, and confidence interval. Utility of alternative study designs may contribute to the USEPA's aims toward

using dose-response modeling, particularly with physiologically relevant models, to assess potential risks of chemical exposure to human health.

Chapter 2

Neurobehavioral Screening Tests

2.1 Screening Tests

Screening tests for neurotoxicity represent the most fundamental level of investigation for forecasting potential neurotoxicity in animal studies (WHO 1986). Screening tests are widely used because they are simple, rapid and economical. A battery of measurements is acquired with these tests that include measurements of behavioral endpoints representing neuro-physiological, neuro-muscular, autonomic and sensorimotor functions. Behavioral endpoints reflect the integration of various functional components of the nervous system and are often used as surrogates for mechanistic processes involved in a subjects response to exposure of a chemical agent. The EPA's testing guidelines developed for the Toxic Substances Control Act (TSCA) and the Federal Insecticide, Fungicide, and Rodenticide Act described the use of neurobehavioral screening tests and established protocols and procedures used in these experimental designs (US EPA, 1991). Since that time the protocols and procedures have been refined and are evolving to become a

standardized first tier screening method for neurotoxicity (US EPA, 1998). These toxicological studies, known as Functional Observational Battery (FOB) tests, are simple to implement, relatively non-invasive and generate behavioral change rapidly.

2.2 IPCS Functional Observational Battery Tests

This thesis utilizes data from a Functional Observational Battery (FOB) as described in Moser et al. (1997a) and implemented in the IPCS' collaborative study. For acute exposure, the study protocol uses 5 dose levels including a control group and 4 testing times (including a baseline measurement) to assess time-related response to chemical exposure. The dose levels and the second testing time were determined via a dose range finding study. Given the cooperative nature of the IPCS sponsored studies several laboratories participated in these studies yielding several independent estimates of dose-response for many of the chemicals tested.

Prior to initiation of the FOB testing a dose range finding study was performed to determine the dosing regime and testing times for a hypothesized time of peak chemical effect (TOPE) (Moser et al., 1997b). The starting dose was chosen based loosely on published estimates of the LD50 (dose which would be lethal to 50% of the subjects). Three doses at constant intervals above and below the starting dose were used for a seven day survival study to determine

the top nonlethal dose to use in the FOB studies. The dose levels used for the FOB designs were then 100%, 50% 25% and 12.5% of the top nonlethal dose. A separate pilot study was then conducted to estimate the TOPE using gait and arousal scores. Thus, the TOPE estimated for gait and arousal scores were taken to represent a hypothesized TOPE for all neurobehavioral endpoints.

Ten animals were randomly assigned to each dose group for the FOB studies. Testing times were established such that a FOB was conducted prior to exposure and at 2-3 subsequent time points after exposure. The first post-exposure testing time was conducted to correspond with the TOPE of the chemical being tested. Subsequent tests were performed at one day and then one week after exposure. Post exposure test times generally did not exceed one week.

The FOB response variables consisted of 25- 30 non-invasive measures designed to assess behavioral alterations with respect to a wide range of neurobiological functions, including sensory, motor and autonomic functions, excitability, neuromuscular strength, and activity level. The entire battery of tests required approximately 6-8 minutes per rat. Assessment of motor activities was used in conjunction with the FOB because of its long history of use for evaluating behavioral effects of chemicals (MacPhail et al., 1989). Motor activity counts represent a broad class of behaviors involving coordinated

participation of sensory, motor, and integrative processes. Neurotoxic agents can lead to either increases or decreases in motor activity counts and organophosphate pesticides such as parathion, the chemical studied in this thesis, have been shown to decrease motor activity counts in neurobehavioral screening studies (USEPA, 1998). The apparatus for motor activity test was left to the discretion of participating laboratories, provided several criteria were met (Moser et al., 1997a). This thesis used parathion motor activity counts to illustrate fitting the biphasic toxico-diffusion model to neurobehavioral screening data.

2.3 Assessment Methodologies

The USF Health Risk Assessment Methodology Group (HRAMG) has used the FOB to develop and test several classes of mathematical models to predict neurobehavioral response to chemical exposure. Liu (2000) developed polynomial models for continuous FOB outcomes; Woodruff (2001) developed a diffusion model that is flexible in describing both transient and persistent nonlinear dose-response relationships seen in the FOB data; Zhu (2001) tested these models on a large number of FOB datasets from both the EPA Superfund and the IPCS studies. More recently Zhu (2005a) developed a class of mathematical models that allow for incorporation of toxicokinetics, and in a series of reports (Zhu et al. 2005a,b) applied these models to the

FOB datasets and illustrated their use in benchmark dose estimation. Common in these statistical analyses was the use of random effects to adjust for biological variation in responses among animals. These models performed well for describing the dose-response patterns observed in the FOB studies. However, physiological relevance of these models relies on simplifying assumptions such as linear or single-compartment kinetics. It would be beneficial to derive models capable of describing exposure related response as well as incorporate parameters that estimate well known physiological phenomena that regulate the time course of a chemicals presence in the body and its affect on the observed response when the system displays nonlinearities such as nonlinear uptake or saturation kinetic processes. Understanding of this process begins with knowledge of how the body processes the chemical (pharmacokinetics/toxicokinetics) and subsequent characterization of how the concentration of the chemical affects the organism of study (pharmacodynamics/toxicodynamics).

Chapter 3

Pharmacokinetics and Pharmacodynamics

In this thesis, references are made to pharmacokinetic and pharmacodynamic (PK/PD) modeling techniques to develop a background on which the generalized biphasic toxico diffusion model was based. It should be noted that while the terms pharmacokinetics and toxicokinetics have essentially parallel meaning in their respective fields, in the strictest sense the former term should be restricted to the field of pharmacology. Since much of the literature devoted to kinetic and dynamic properties of chemical exposure has arisen from the field of pharmacology, we define the modeling approach using the PK/PD modeling paradigm and apply the biphasic model to a toxicological study.

3.1 Compartmental Pharmacokinetics

Orthodox pharmacokinetic and toxicokinetic studies deal with changes of drug concentrations in the plasma or an organ over time in an attempt to describe how the body absorbs, distributes and eliminates a drug (Torda et al. 1994). The compartmental approach to pharmacokinetic estimation views the body as being composed of a

number of pharmacokinetically distinct compartments. Each compartment can be thought of as an imaginary space in the body representing a combination of various tissues and organs, among which the drug interacts. Anatomical composition of the compartment is unknown and in most cases its analysis is often of little value (Kwon 2002). Compartmental models are designed to provide a conceptual understanding of distributional behaviors of a drug between the plasma and other tissues or organs in the body and estimate various pharmacokinetic parameters including plasma concentrations, apparent volumes of distribution and rates governing elimination and clearance. It is recognized that these compartment models, while estimating physiological properties, still represent empirical fits to the data. However, the physiological interpretations are important for extrapolating information from animal studies to regulatory data on human health (US EPA 1998). Consider, for example, the simplest of pharmacokinetic models describing IV bolus injection of a substance into the circulatory system and subsequent elimination (K_e) from the body (Figure 1).

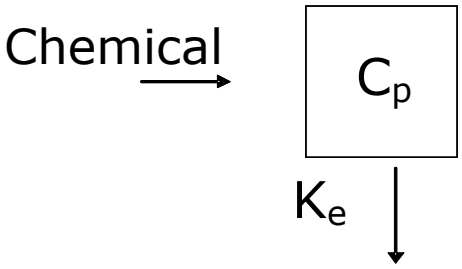


Figure 1. Illustration of simple one compartment pharmacokinetic model

Concentration of the chemical in the compartment is highest immediately after injection and concentration at any time *t* is governed by a single exponential rate of elimination: $C_p(t) = C_0 * e^{(-K_e * t)}$

Where;
$C_p(t)$ = Concentration in the circulatory system at time *t*
C_0 = Concentration of chemical immediately after IV bolus injection
$e^{(-K_e * t)}$ = an exponential elimination rate constant.

This first order, mono-exponential function is typically used to model plasma concentration versus time curves with direct injection into a central compartment yielding estimates of compartment volume and chemical elimination. One compartment behavior of plasma concentration does not necessarily imply that the chemical is at the same concentration in all the tissues and organs in the body. Rather,

this implies that the concentrations are in instantaneous equilibrium with those in the plasma upon drug administration.

An expansion of the one-compartmental approach to pharmacokinetic modeling is to allow for the distribution of the chemical from the plasma compartment into a peripheral compartment (brain, muscle tissue, etc). When distribution of a chemical from the plasma into certain organs or tissues is substantially different from the central compartment, multi-compartment models allow for the incorporation of one, or several, peripheral compartments (Figure 2).

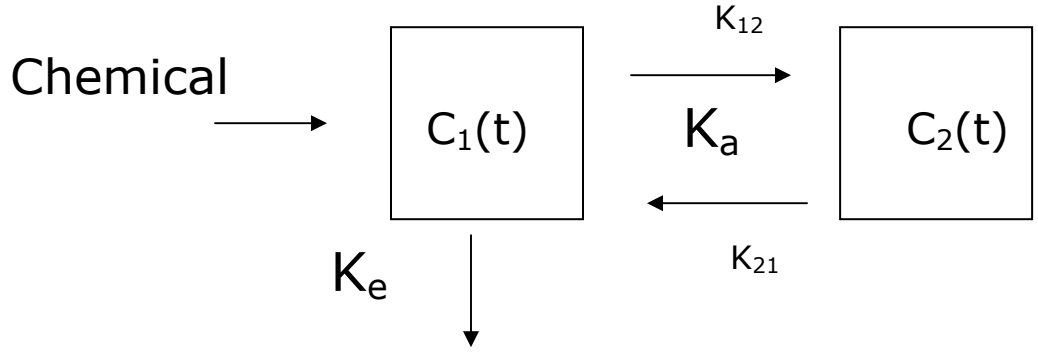


Figure 2. Two compartment PK model describing the pathway of chemical introduction and elimination from an organism

The two compartment model is typically defined where the administered chemical is delivered and eliminated from the central compartment and distribution between the central and peripheral compartments is controlled via two distribution micro-constants (K_{12}

and K_{21}) characterized by the distribution constant K_a (Gabrielsson and Weiner 2000). When these rate constants reach a steady state, then the distribution of the chemical reaches equilibrium and the mechanism governing concentration in the central compartment is controlled through the elimination rate K_e . Methods of estimating the concentration in the central compartment via a two compartment paradigm include the bi-exponential function

$$C_p(t)=\theta_1 e^{-K_a*t} + \theta_2 e^{-K_e*t}$$

where K_a represents the distribution phase from the plasma that includes absorption into the second compartment and K_e represents the post equilibrium or terminal elimination phase governed by elimination of the chemical from the central compartment. Chemical concentration in the plasma is highest immediately after injection in the two-compartment model and behaves similarly to the one compartment model except initially when absorption into the second compartment causes an accelerated depletion of the chemical concentration from the central compartment (Figure 3).

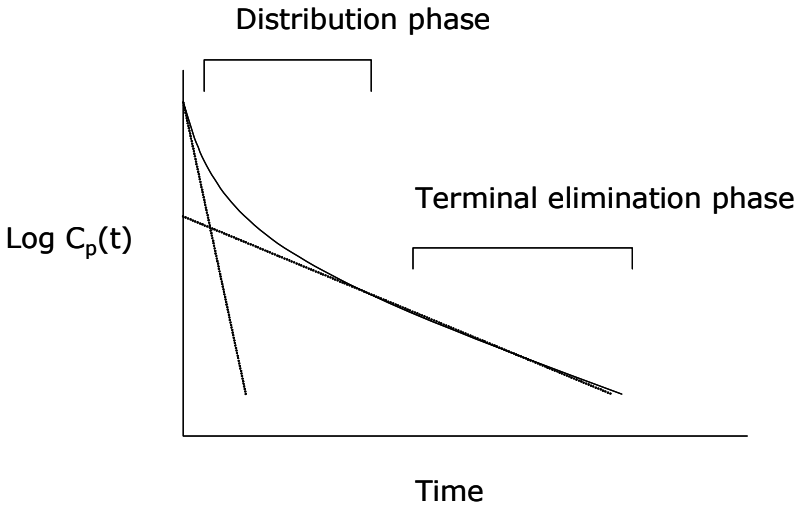


Figure 3. Bi-exponential decline of concentration in the central compartment after iv bolus injection when chemical distribution can be described using two-compartment model (Taken from Kwon 2002).

3.2 Nonlinear Kinetics

Any pharmacokinetic process of a chemical (i.e. absorption, distribution, metabolism or excretion) that cannot be described with first order (linear) kinetics can be considered nonlinear kinetics. Nonlinear kinetics implies deviations in the rate of change in chemical concentration from first order kinetics in a manner that is dose and /or time dependent. Dose-dependent nonlinearity can be due to any carrier mediated process such as metabolism or active transport that displays transient saturation or cooperativity at high concentrations (Kwon 2002). Dose-dependent kinetics may be observed as a

stimulated or suppressed response function where increases in chemical concentration result in non additive increases in concentrations in peripheral compartments in either a synergistic or capacity limited manner.

3.3 Pharmacodynamics

To this point we have considered the relationship between administration of chemical into an animal and some of the basic pharmacokinetic parameters describing the time course of chemical concentration in biological fluids. Obviously, there are many other factors, known and unknown that govern the specific "effects" of chemical exposure on the exposed subject. The study of Pharmacodynamics (PD) is designed to characterize the effect of the chemical on the body. If concentration in the plasma and effect site is in rapid equilibrium, PD models may adequately serve to estimate the pharmacological effects of chemical in the body. These models are valuable in capturing the nonlinear aspects of response patterns often seen in pharmacological experiments and in estimating various pharmacodynamic parameters used for establishing dosing regimens in clinical studies (Gabrielsson and Weiner 2000). However, advancements in the field of pharmacology and toxicology have included the realization that more often in *in vivo* experiments,

pharmacological effects take time to develop.

Pharmacokinetic/pharmacodynamic (PK/PD) models attempt to link concentration in the plasma and response to chemical exposure by linking the pharmacokinetic properties of a chemical to the biological response observed at the effect site. While direct and simultaneous measurement of chemical concentration at the effect site and its pharmacological effect is the most desirable approach to reveal the true pharmacodynamic profiles of a substance, it is seldom feasible to measure chemical concentration at the effect site because of limited accessibility and availability to the site (Kwon, 2002).

3.4 Pharmacokinetic/Pharmacodynamic Modeling

Pharmacokinetic/Pharmacodynamic modeling has become a common mechanism for elucidating the chemical/effect relationship when a distributional delay exists or when the system is subject to time or dose-dependent kinetic and/or dynamic changes. Effect compartment models are a class of PK/PD models used to account for differences in concentration between the plasma and effect compartments by introducing an equilibrium rate constant K_{eo} (Holford and Sheiner, 1981) (Figure 4).

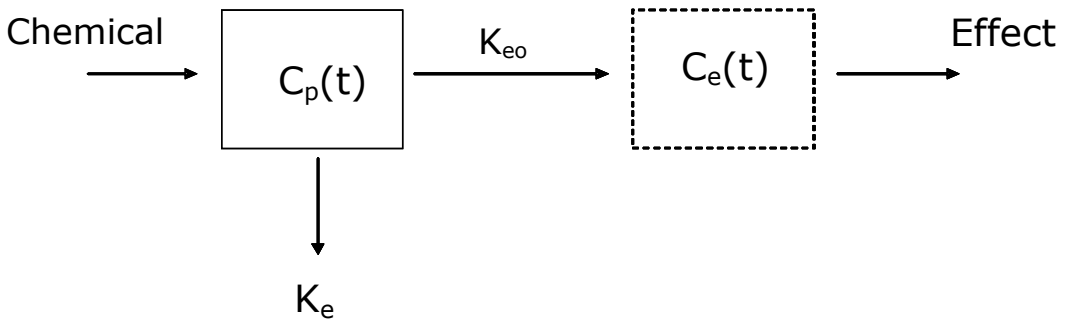


Figure 4. Illustration of the Effect-Compartment model.

This rate constant K_{eo} eliminates the discrepancies observed between the plasma and effect compartments resulting in an estimate of chemical concentration (C_e) in the effect compartment. For example, an equation used to link a single compartment pharmacokinetic equation assuming intravenous injection of drug into the central compartment to a pharmacodynamic model is the following biphasic kinetic equation.

$$C_e(t) = \frac{D}{V}\frac{K_{eo}}{K_{eo} - K_e}(e^{-K_e t} - e^{-K_{eo} t})$$

Where:

$C_e(t)$ = Concentration in the effect site at time t

$\frac{D}{V}$ = dose/volume equivalent to Ce(t=0)

K_e = Elimination rate for the central compartment

K_{eo} = Distribution rate for the effect compartment

This model is used to account for differences between the central and effect compartments caused by distributional delay (Kwon 2002). This function becomes central to the biphasic toxico-diffusion model discussed in Chapter 5.

Once $C_e(t)$ is derived, effect compartment models can then be used to estimate pharmacodynamic response. Since observed response can be represented by either an elevation of the measured effect or an inhibition of the effect, these effect compartment models are classified as either E_{max} (elevated response) or I_{max} (inhibited response) models.

$$E(t) = \frac{E_{\max} C_e(t)}{EC_{50} + C_e(t)}$$

A baseline response E_0 can be added to reflect change from baseline. The term E_{max} in the equation corresponds to the theoretical maximum effect while EC_{50} represents the "half-life" time at which the effect is half of the maximum response. A power term (n) can be added to the E_{max} or I_{max} models making them sigmoid models (sometimes referred to as the Hill equation) to represent sigmoid concentration versus effect curves related to carrier mediated processes such as enzyme cooperativity (negative or positive) affecting the concentration effect curve.

$$E(t) = \frac{E_{\max} C_e^n(t)}{EC^n{}_{50} + C_e^n(t)}$$

Exponents less than 1 increase the initial slope of the curve while an exponent greater than 1 makes the curve more shallow initially. These curves converge around the EC_{50} value and then cross with larger exponents reaching E_{max} more quickly while smaller exponents yield asymptotic values less than E_{max}. An exponent of 1 reduces the equation to the standard hyperbolic E_{max} model.

3.5 Michaelis-Menten Equation

The effect compartment models are based on a well defined equation developed in the early twentieth century to describe the fermentation of cane sugar (i.e. sucrose) via hydrolysis into glucose and fructose. Michaelis and Menten derived an equation built upon earlier work by Henri and others by carrying out definitive experiments using the enzyme invertase (Cornish-Bowden 1995). They found that the rate of the reaction *v* was dependent on the substrate (sucrose) concentration (a) and limited by the concentration of enzyme in the reaction. The reaction is described by the equation:

$$v = \frac{V_{max} a}{K_m + a}$$

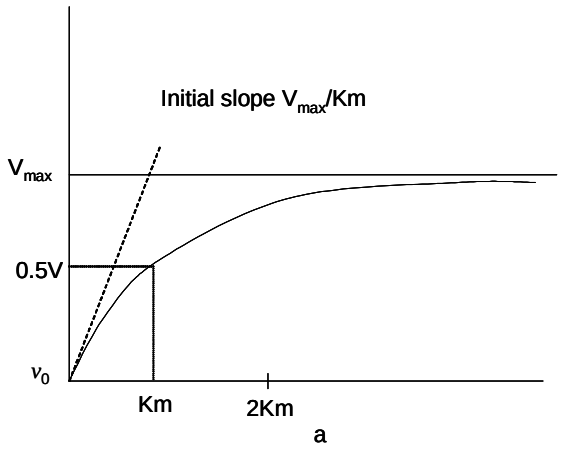


Figure 5. The Michaelis-Menten equation with illustration of the relationship between substrate concentration (a) and velocity of the reaction. Km is the concentration of substrate that equaled half the maximum velocity of the reaction.

Where: V_{max} is the limiting rate for the velocity of the reaction and K_m is defined as the Michaelis constant describing the substrate concentration at which the velocity of the reaction is 1/2 that of V_{max}. Among the many important contributions to the description of this reaction was the measurement of the initial rate of the reaction (V_{max}/K_m) at different sucrose concentrations thereby avoiding complicating factor such as reverse reaction, product inhibition and inactivation of the enzyme (Cornish-Bowden, 1995). This equation is recognized as the fundamental equation of enzyme kinetics and has been widely used to describe biological processes in areas outside of its original intent. In pharmacological studies, a relationship described

by this type of equation is referred to as having Michaelis-Menten kinetics.

Chapter 4

Nonlinear Mixed Effects Models

The HRAMG has advanced the estimation of dose-time-response in FOB studies by using non-linear mixed effects methods for estimating response to chemical exposure. These models capture the dose-dependent response to chemical exposure while allowing for individual subject variation in their natural responses to the measurement instrument in the absence of chemical exposure. Increasingly, these nonlinear models are evolving to incorporate parameters thought to describe well known physiological mechanisms governing response.

4.1 Toxico-Diffusion Model

Zhu (2005) used a re-parameterized version of the Michaelis-Menten equation for the analysis of the FOB data.

$$f(d,t) = \frac{B * conc(d,t) * t}{1 + C * conc(d,t) * t}$$

In this equation

$$conc(d,t) = \frac{d}{V}\exp(-K_e t)$$

is dependent on the parameter K_e which can be interpreted as the elimination rate under intravascular administration (Zhu 2005a). The parameter *V* represents the volume of the circulation system or central compartment. In the toxico diffusion function, *V* is absorbed into the parameters *B* and *C* in the function.

This function resembles the Michaelis-Menten equation, but with concentration-dependent coefficients *B*conc(d,t)* and *C*conc(d,t).* The coefficient *C* can be negative as long as the denominator is positive within the experimental range (Zhu 2005a). By incorporating a baseline response (*A*), the equation describes the change in response from the initial condition measured prior to exposure throughout the time course of the study. Assuming that chemical concentration is directly linked to response, rapid elimination of the chemical corresponds the function *f(d,t)* quickly reaching a peak value at the time of peak effect (TOPE), and returning tobaseline. However, if the compound remains in the subject's system, *f(d,t)* may not return to the baseline line level, characterizing persistent dose effects.

This function, relying on a one compartment kinetic paradigm, results in a dose independent TOPE. As *t* varies from 0 toward infinity, *f(t,d)* varies from baseline to a maximum at $t = 1/K_e$, then back to baseline, irrespective of dose level. Statistical evidence of neurotoxic

effects is present when either the coefficient *B* or *C* is statistically non-zero.

4.2 Parameter Estimation Methods

Implementation of the toxico-diffusion function represents a situation where the observed response is predicted based on a nonlinear function of the estimated parameters in the model. Often in application of these nonlinear models it is advantageous to incorporate a random effects parameter in addition to the fixed effects coefficients to account for natural variation in biological response to the testing instrument. Thus, random effects represent deviations from the population average and can enhance parameter estimation and hypothesis testing procedures by accounting for a source of variation in the data otherwise subjected to the error term. Given the short temporal sequences of FOB data, it is rare that the data can accommodate more than one random effect even if biologically feasible (Zhu 2005a). Models that incorporate both fixed and random components are often termed "mixed-effects" models and a detailed description of their development and implementation can be found in Pinhiero and Bates (2000). These models generally use maximum likelihood (ML) methods for parameter estimation. Likelihood functions for mixed-effects models are generally complex, and closed form solutions are generally not available. Implementation of ML requires

iterative numerical procedures such as Newton-Raphson algorithm, the EM-algorithm, and more often the combinations of them (Lindstrom and Bates 1988). The model fitting techniques used in this thesis rely on the Newton-Raphson algorithm.

4.3 Model Selection and Diagnostics

There are several important assumptions associated with mixed effects models that require validation. Namely, random effects are assumed to be normally distributed around a mean of zero; the random effects and the error term are assumed to be uncorrelated, and the variance is assumed constant across different experimental conditions. Choosing the most appropriate model to represent the response observed in the data requires both objective criteria and sound investigative principle. The Likelihood Ratio Test (LRT) and Information Criteria (AIC and BIC: Pinhiero and Bates, 2000) are useful tools for guiding appropriate model selection. Significance testing using the LRT relies on the Chi-squared distribution (i.e. $2(\log(L_1)-\log(L_2)) \sim \chi^2_p$) , where L_1 is likelihood function of the expanded model and L_2 is the existing model, and p is the number of additional parameters in L_1 . While this statistic is readily available in most computer software packages, the one-sided alternative test may require weight adjustments in some cases (Zhu, 2005). Another

restrictive requirement of LRT is that one model must be a sub-model of the other comparison model. Information Criteria (Akaike, 1973) are a generalization to the Likelihood Ratio Test. They add to the likelihood ratio a term to penalize the inclusion of excessive terms in the model. They do not require one model being a sub-model of the other, but do require the models follow the same family of distributions. Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) are two popular methods that yield values for comparing the goodness of fit of two models. The model with the smaller value of AIC and/or BIC is favored. Investigative principles include graphical comparisons of the fitted model with the raw data and residual plots to check for randomness of error terms. These are helpful tools for selecting the appropriate model in conjunction with statistical criteria such as the Log-Likelihood Ratio Test (LRT) and Information Criteria (AIC and BIC). Graphical tools are useful for visualizing not only the shape of the dose-response but also the agreement of the predicted and observed responses. Residual plots check the assumptions of independent, normally distributed errors and identify potential outlying observations.

Chapter Five

Expanding the Toxico-Diffusion Function

5.1 Biphasic Toxico Diffusion Function

This thesis generalizes the toxico-diffusion function developed by Zhu (2005) to estimate the dose-time-response relationship in which a dose-dependent TOPE is observed. The biphasic toxico-diffusion function is a modification of the biphasic kinetic equation discussed in chapter 3 that accounts for discrepancies between the central and effect compartments in pharmacokinetics. Its foundation is the toxico-diffusion function with the addition of a second exponential term to achieve a dose-dependent TOPE. The rate of the second exponential term is dependent on dose as well as time. This second exponential term forms a complex exponent with the elimination rate either in the numerator (Equation A) to describe accelerated TOPE or in the denominator (Equation B) to describe the delayed TOPE often observed in capacity limited kinetics.

$$f(t,d) = \frac{Btd * (\exp(-K_e * t) - \exp(-K_{e0} * dose * t))}{1 + Ctd * \exp(-K_e * t)} \quad \text{(A)}$$

$$f(t,d) = \frac{Btd * \exp(-K_e * t)}{1 + Ctd * (\exp(-K_e * t) - \exp(-K_{e0} * dose * t))} \quad \text{(B)}$$

Therefore, these models do require some *a priori* knowledge of the dose-response relationship to select the appropriate model to describe dose dependent TOPE. At some point in the time course of chemical exposure, assuming $K_{eo} >> K_e$, the second exponential term of the biphasic model reduces to zero and the model reverts to the standard toxico-diffusion function.

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

As with the pharmacodynamic models, the biphasic toxico-diffusion models are not intended to be mechanistic in the literal sense of describing actual pharmacologic or kinetic processes. Rather, here the purpose of these models is to incorporate pharmacokinetic concepts into a mathematical model that allows for a dose-related shift in TOPE that may be related to well known pharmacokinetic processes such as nonlinear disposition, tissue or protein binding, metabolism or clearance kinetics observed in pharmacological and toxicological studies (Gabrielsson and Weiner 2000). Irrespective of the actual mechanism governing the observed shift in TOPE, these proposed

models may be useful for describing the FOB data and elucidating dose-response relationships which exhibit dose-dependent TOPE.

To illustrate the use of the biphasic toxico-diffusion models, parathion motor activity counts were chosen as the response variable of interest due to an observed shift in the time of maximal response at higher dose levels. In the IPCS studies parathion was introduced by oral gavage in corn oil solution at doses of 0.85, 1.69, 3.38 and 6.75 mg/kg. Parathion is readily absorbed through the digestive tract and detection in the bloodstream has been reported immediately after oral administration (INCHEM 2004). Nonlinear binding of the toxic parathion metabolite, paraoxon, has been reported for red blood cells (with supra-linear dose-response) and brain tissue (with a sub-linear dose-response) (Vogel et al. 2002). Kramer et al. (2002) used a three compartment model to describe the pharmacokinetics of Methyl Parathion but failed to elucidate the pharmacodynamic properties associated with this hazardous organophosphorus pesticide. For the model described in this thesis to be related to physiological processes, bioavailability is assumed to be 100% and absorption from the gastrointestinal tract is assumed to be nearly instantaneous.

5.2 Parathion Activity Counts

Examination of the raw data plots reveal the relationship between dose, time and activity counts observed after a single exposure to parathion (Figure 6). These grouped data plots are constructed such that the control group is the first graph (starting on the left), followed by panels of increasing dose to the right. The x axis is in log scale for display purposes only.

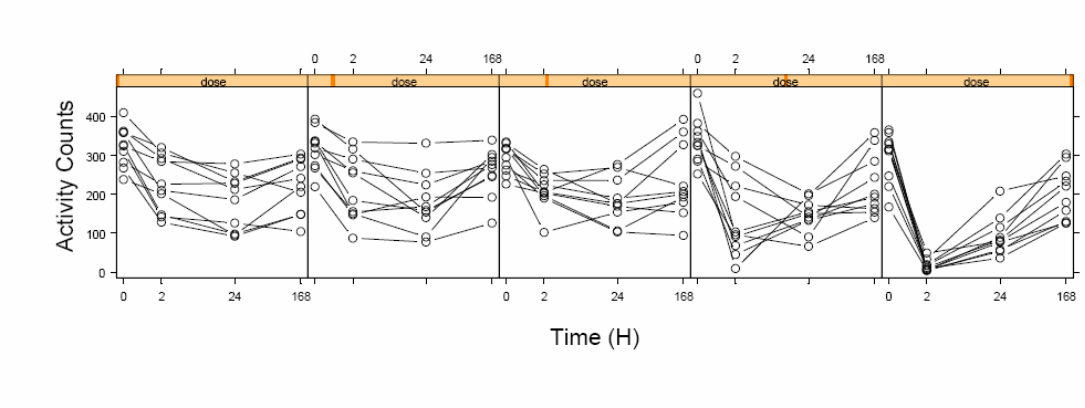


Figure 6. Observed response in activity counts of rats subjected to acute parathion exposure. Note: Panels are arranged in an increasing order of dose (0, 0.85, 1.68, 3.38, 6.75) from left to right.

It is apparent that dose increases resulted in changes of the response trajectories (Table 1). The two highest dose groups (3.38mg/kg and 6.75 mg/kg) appeared to have vastly lower activity counts than the lower dose groups. The observed TOPE associated with the higher dose groups was 2H compared with a TOPE at 24H for the controls and the lowest dose group.

Table 1. Dose group average activity counts for motor activity counts after acute parathion exposure with standard deviation in parenthesis

Dose	Time			
	0	2	24	168
0	323.6 (51.3)	225.6 (72.1)	180.2 (71.4)	222.3 (70.0)
0.85	316.0 (53.2)	217.8 (83.9)	178.9 (76.0)	257.5 (60.5)
1.69	291.9 (37.5)	207.6 (44.5)	185.5 (59.9)	232.5 (95.8)
3.38	336.2 (57.8)	139.6 (99.2)	144.0 (42.4)	224.9 (77.9)
6.75	296.2 (64.1)	15.9 (14.7)	93.6 (49.9)	201.7 (67.7)

These aspects of the observed data suggested that the parathion activity count data may allow for an application of the biphasic numerator model. Therefore, the biphasic numerator model (equation 1) was applied the parathion motor activity counts to describe the apparent acceleration of TOPE observed in the data.

5.3 Fitted Response

The fitted biphasic model predicted a decline in activity counts after exposure to parathion, and the predicted trend resembles the observed pattern. However, only the intercept A and the rate parameter K_e were statistically significant, the slope factors B and C and K_{eo} are insignificant (Table 2). A large amount of variation in the data was attributable to the between rat variation at baseline and was accounted for by a random effect for baseline activity count. The random effects had a standard deviation 35.9, about 54% of residual standard deviation (Table 2).

Table 2. Summary of biphasic model fit to parathion activity counts

Parameter	Estimate	Standard Error	DF	t-value	p-value
A	258.0129	7.91744	146	32.588	<.0001
B	-72.3701	83.48372	146	-0.867	0.387
C	0.2265	0.35207	146	0.643	0.521
K_e	0.1400	0.03561	146	3.931	0.0001
K_{eo}	0.3198	0.21072	146	1.518	0.131
Random Effects	Intercept	Residual			
	35.93553	65.08784			

The fitted biphasic model appeared to be a reasonable fit the observed data. The fitted dose-response pattern resembles the observed data in aspects of TOPE and magnitude of response changes.

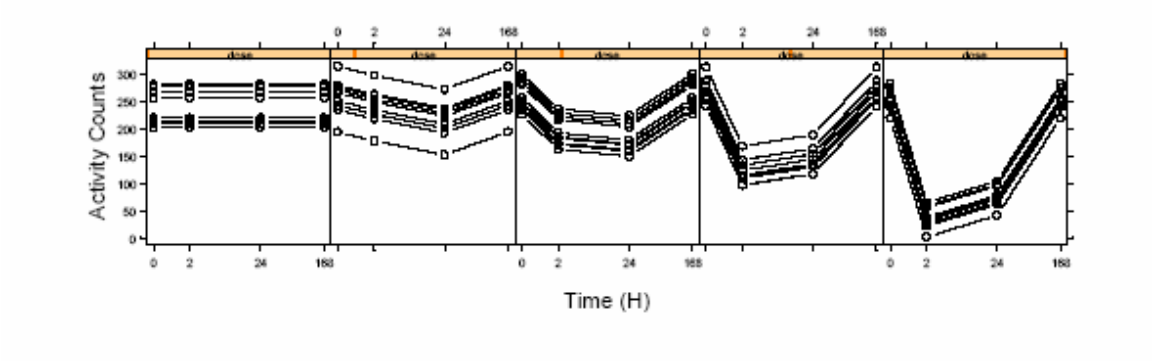


Figure 7. Fitted response of biphasic numerator model for rats subjected to acute parathion exposure.

Evidence of a good fit of the model to the data comes through examination of the residual plots. Residual plots express the difference between the observed and predicted responses, termed “error”. An assumption of nonlinear mixed effects models is that the errors are

independent and normally distributed. Standardized residual plots for the parathion activity counts (Figure 8) suggest that the errors are centered on a mean of zero and fairly randomly distributed.

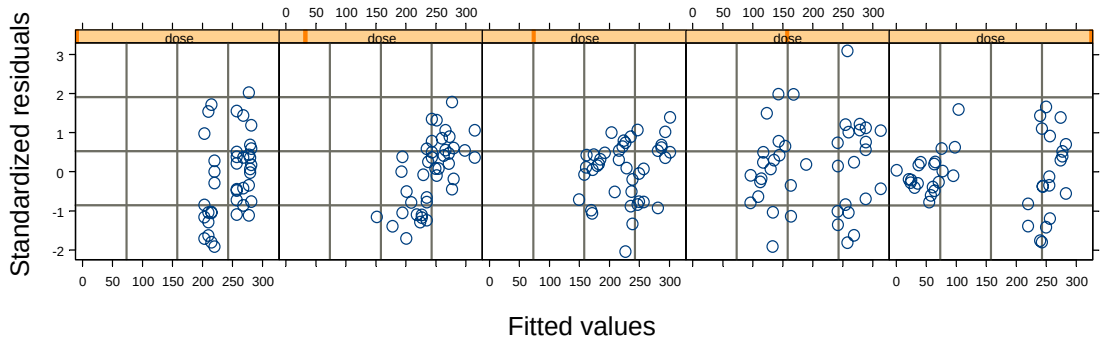


Figure 8. Residual plots of biphasic model for rats subjected to acute parathion exposure. Note: Panels are arranged in an increasing order of dose (0, 0.85, 1.68, 3.38, 6.75) from left to right.

The NLME models further assume that the errors are constant across dose group while often in biological settings the variance will increase in proportion to the mean response. When there is evidence of non-constant standard deviation, NMLE can incorporate a separate standard deviation to each dose group and verify the resulting model improvement using the Likelihood Ratio Test (LRT). The addition of dose group specific standard deviations to compensate for heteroscedascity did not significantly improve the model fit according to the likelihood ratio test (log(L2)-log(L1)) value of 7.28 and a corresponding p-value of 0.122. This suggests that a constant standard deviation across dose groups is acceptable.

5.4 Dose-Dependent TOPE

The fitted model also reveals different times of peak effect determined by dose level (Figure 9): the higher dose groups clearly had an accelerated TOPE. The TOPE varied from 11.5 hours for the lowest dose group (0.85mg/kg) to 7.3 and 7.1 hours for the 3.38mg/kg and 6.75mg/kg dose groups, respectively.

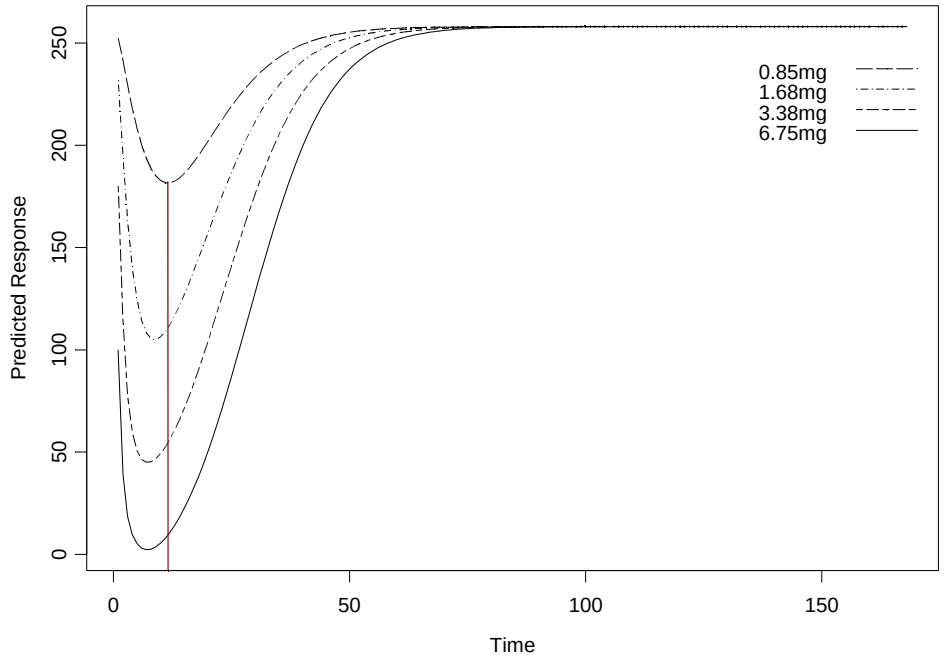


Figure 9. Dose specific trajectories for biphasic model on rats subjected to acute parathion exposure.

The fitted model can also be examined for each subject after accounting for the random effect at baseline (Figure 10). Indeed, the predicted model appears to capture individual (red) deviation when compared to the population average (blue).

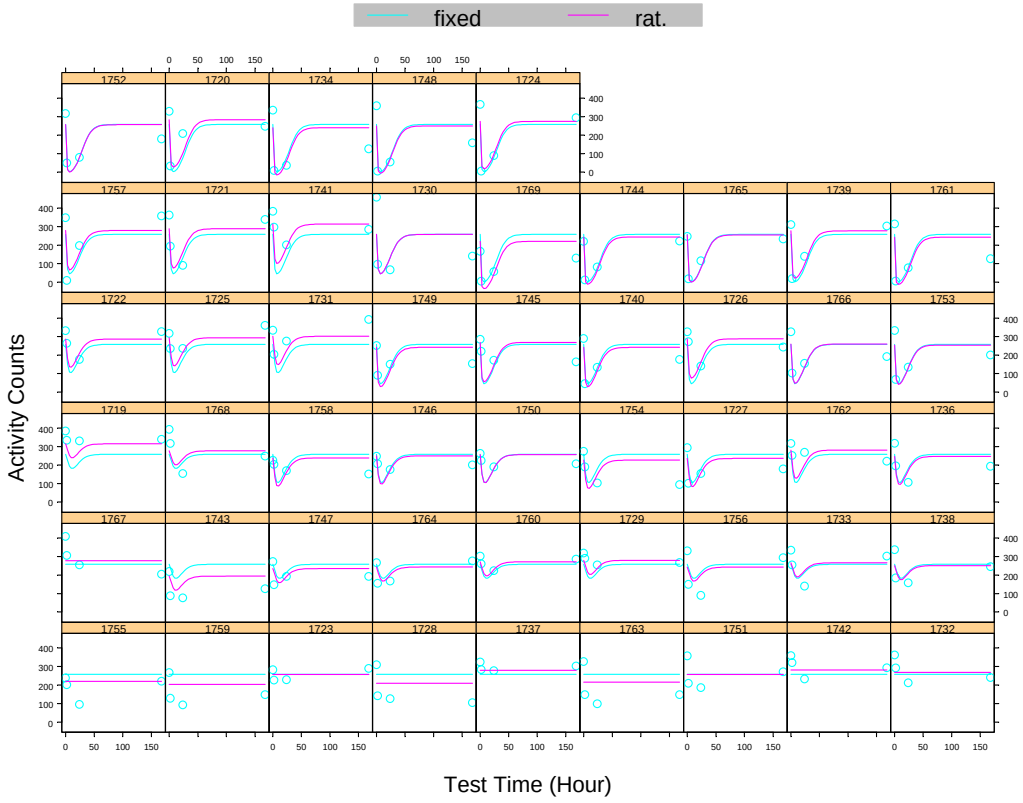


Figure 10. Individual specific trajectories for biphasic model on rats subjected to acute parathion exposure.

The biphasic model can be viewed as a generalization of the toxico-diffusion model to allow for a dose-dependent TOPE. Given that these models have this hierarchical structure, Information Criteria and the Log Likelihood Ratio test were used to test for improvement in the fit of the biphasic model relative to the fit of the toxico diffusion model (Figure 11). Both Information Criteria and the Log Likelihood Ratio test suggest a significant improvement of the biphasic model overall (Table 3). Despite the fact that the dose-dependent coefficients (B and C)

and the second kinetic coefficient K_{eo} in the biphasic model are statistically insignificant, due perhaps to limited number of data points (in time or subjects) in the experiment, the biphasic model is attractive because it permits dose-varying TOPE as indicated by the motor activity count data. Further, the magnitude of response predicted by the biphasic model was dramatically different between the two models with the toxico-diffusion model predicting peak response (in this case lowest activity count) of 100 at the highest dose level while the biphasic model predicted a peak response (lowest activity count) of three; the true response observed in the data (see figure 6). Correctly predicting the magnitude of peak effect is important as the peak effect could have dramatic implications on benchmark dose estimation. On the grounds of toxicology, it is generally accepted that the TOPE is less likely to be a constant. With only a limited number of time points (4 in the case of motor activity counts) spread over a wide range, however, variance in TOPE is less likely to be detectable from the FOB data. Therefore, verification of a true dose-related TOPE is difficult because the statistical power in detecting the variation in observed TOPE from standard FOB assay data is low.

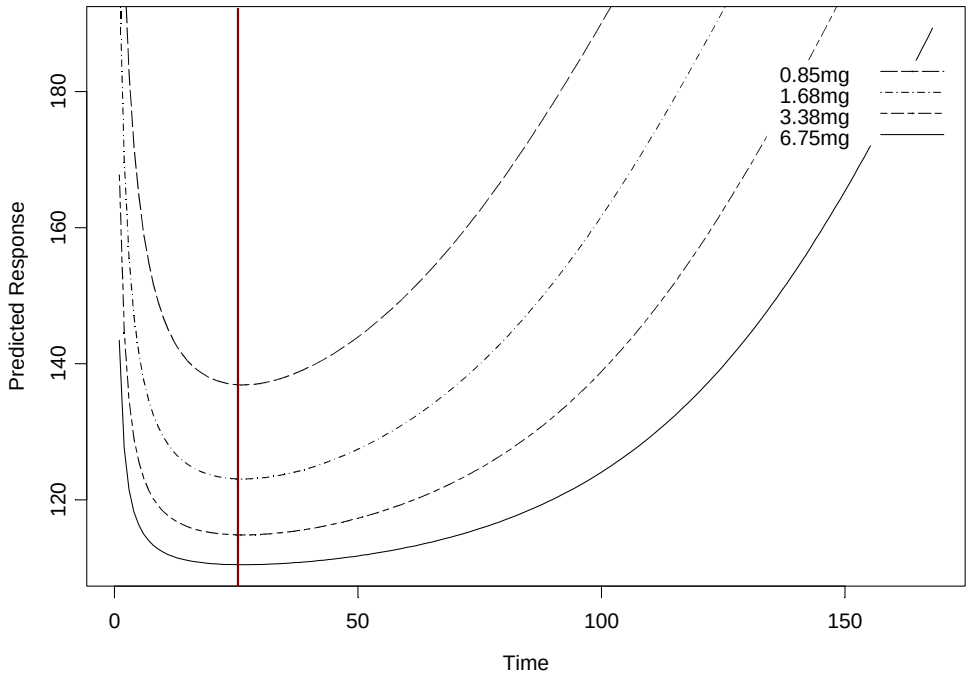


Figure 11. Predicted response to parathion by the toxico-diffusion model.

Table 3. Results of log likelihood (LL) ratio test comparing the toxico-diffusion (TD) and Biphasic Numerator (BNM) models.

Model	DF	AIC	BIC	LL	Chi Sq.	Pr>Chi.Sq
TD	6	2297.751	2317.541	-1142.876		
BNM	7	2291.730	2314.818	-1138.865	8.021	0.0046

The analysis thus far has demonstrated that FOB assays using scarcely spaced experimental time points will likely fail to generate sufficient data for dose-response modeling especially when dose varying TOPE's are considered. The next chapter focuses on considering FOB designs that may provide enough information about the true dose-response necessary to capture dose-dependent response

patterns of neuro-toxicological effects, particularly dose-dependent time of peak effects, in a valid and reliable manner. In the next chapter, computer simulation is used to explore the most effective FOB design protocols to predicted dose-dependent TOPE using the biphasic toxico diffusion model.

Chapter 6

Simulation

6.1 Simulation Rationale

While the biphasic toxico-diffusion model appeared advantageous relative to the toxico diffusion model in permitting dose-dependent TOPE, the parathion motor activity count data leaves some uncertainty about the model as the parameter estimates for the B, C and K_{eo} were statistically insignificant. This lack of statistical power seemed to result from insufficient data generated under the current FOB design. Specifically, the sparse spacing of experimental testing times is believed to provide insufficient data for model parameterization. Therefore, Monte Carlo simulation was adopted to investigate the design protocols necessary to recover the key characteristics of dose-dependent response along the time course as predicted by the biphasic model. The primary objective of the simulation was to investigate the benefits of considering alternative time spacing and/or adding additional experimental testing times and/or experimental subjects to the FOB design to recover the key characteristics of the biphasic model with regard to parameter

estimation. Assuming that the biphasic model represents the true underlying response of rat motor activity to acute parathion exposure, we could construct a dataset through simulation that contained the underlying response at a sequence of possible testing times in addition to the experimental times in the FOB study. We could also control the number of subjects in each dose group. The simulation experiments then allowed us to evaluate, empirically, the efficiency of designs with various locations and frequencies of testing times as well as with additional subjects needed for the biphasic toxico-diffusion model to capture the dose and time dependent characteristics of the "true" response.

6.2 Simulation Methods

We obtained the parameter estimates from fitting the biphasic model to the motor activity count data of rats exposed to parathion. These estimates were then used as the population ("true") parameters in the biphasic function to define the "true" dose-response model. Dose levels remained the same as those used in the original parathion experiment. Within the content of simulation, the fitted model represents the true underlying mean response to parathion, and the random effects and random errors govern variation among rats in a given population. Generating simulation datasets required several steps:

Step 1. Generate the mean underlying dose-time-response curve using the biphasic model (with the parameter estimates in Table 1) at selected dose and time levels. This step results in average response specific to the given dose and time involved in the designed experiment (Figure 12).

Step 2. Add random effects. To account for each subject's deviation from the average score at baseline, random effects were generated for each subject using a normal distribution, $N(0, \sigma D)$, where the standard deviation (σD) was taken from the estimate (35.94) of the fitted biphasic response model. The random effects were added to the mean response at every dose-time point in the form of the random intercepts. In the present simulation there was one random effect for each rat.

Step 3. Add random errors. Random errors were generated for each individual observation from a normal distribution N(0, σ_{ε}) and added to the mean response to represent measurement variation (σ_{ε} = 65.09) around the mean response.

Step 4. Repeat Steps 1-3 n times to generate data of n rats per dose group. This process generates one simulated experiment.

Step 5. Repeat Step 4 one thousand times to generate 1000 replications of the experiment.

Based on this simulation protocol, we generated motor activity scores for n rats in each dose group at a sequence of desired testing times. We simulated responses at each of the time points used in the original study (i.e. 0,2,24,168) as well as additional points in 4 hour intervals between times 4 and 24 (i.e. hours 4, 8,12, 16,20). In this way we had information at a multitude of testing times around the time of peak effect, which allowed us to identify effective designs for recovering key characteristics of the biphasic model.

The purpose of analyzing these simulated experiments was to assess the efficiency of these experiments with respect to recovering the "true" underlying dose-response information, specifically dose-dependent TOPE, with a high level of statistical certainty (power). Thus, we first considered two groups of simulation experiments, 4-time point designs and 5-time point designs with ten subjects per dose group. While a larger number of time points are statistically desirable, designs with 4-5 time points are more practical and more closely follow the EPA neurotoxicity risk guideline (USEPA, 1998). Intuitively, a time point in the neighborhood of the true TOPE would provide the most relevant information on the TOPE. Therefore, in the 4-point design, we retained three time points, 0, 24, and 168 hours from the original design protocol and let the 2^{nd} time point vary between 2 h and 24 h. These design variations aim at assessing the efficiency of

the original FOB experimental design. In the 5-point design, an additional time point between 4 and 24 hour was added to the original FOB experiment.

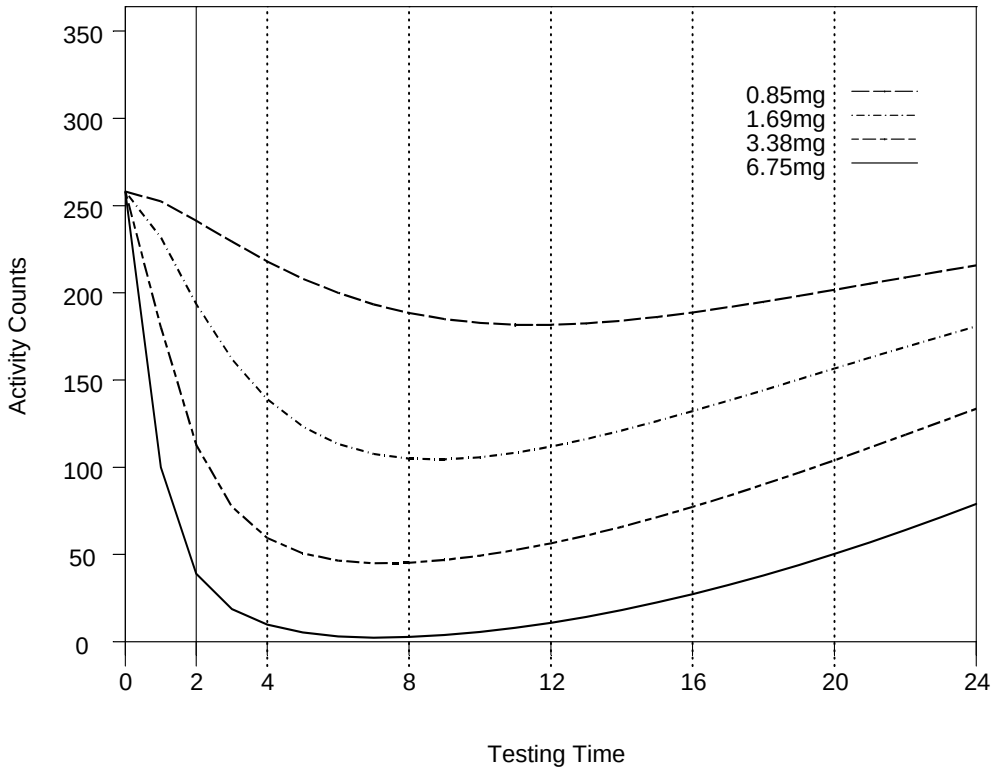


Figure 12. Theoretical "true" dose-dependent biphasic model response between 0 hours and 24 hours. Candidate experimental testing times used for simulation are indicated by the broken vertical lines

In addition to the 4-point and 5-point designs, we also considered designs with 6, 7, 8, or even 9 time points for the purpose of statistical dose-response modeling. These simulated experiments helped illustrate the relative gain of using additional points in

increasing statistical power or gaining information on the dose-response. In order to see the effects of sample size on statistical power, we also considered using more subjects (i.e. 20, 30) at each dose level under the 4 and 5- time point designs. Note there are three 5-point designs that did not use 2 hours as an experimental point but rather contain 2 testing times between 4 and 12 hours as well as at hours 0, 24 and 168.

One thousand data replicates of each of the 35 design trials listed in Table 4 were generated and analyzed to evaluate each design. Efficiency of the design was evaluated in the following ways:

1) The convergence rate - the number of fits on the biphasic numerator model that reached successful model convergence out of the 1000 replicate trials.
2) Statistical power: The percentage of times the t statistic for each parameter exceeded its 95^{th}-percentile (p-value less than 0.05, or type I error at 0.05).
3) Bias = ($\Sigma(\hat{\phi}_i - \phi)$/n): The difference between the true model parameter and the average of 1000 replications of the parameter estimate.

4) Mean Squared Error = $(\Sigma(\hat{\phi}_i - \phi)/n)^2 + \sigma^2$: The Bias squared + the sample variance of the 1000 replications of the parameter estimate.

5) The Bias and MSE of the Time of Peak Effect (TOPE).

These statistical criteria were used to compare the designs and identify the more efficient designs with respect to dose-response modeling of the biphasic toxico-diffusion model. It should be noted that the assessment of power for any parameter estimate precludes the situation that the parameter is not meaningful both biologically and mathematically. This is because any parameter can be imposed and signified statistically with a sufficiently large dataset. However, the case of biphasic model indeed precludes the possibility of statistical manifestation. While found to be statistically insignificant in the original model fit, the model parameters have inherent importance in the model function to describe the response (e.g. initial rate of change in response, "half-peak" effect dose, and dose dependence in TOPE). Comparison with different models suggested the magnitude of the parameter estimates were non-trivial, but the design was underpowered to ascertain the value. Within the context of simulation, the scenario of such a statistical manifestation can be plainly excluded

because data were generated from the biphasic model and all involved parameters are mathematical quantities in existence.

Table 4. Simulated experimental FOB designs using 10, 20, and 30 subjects per dose group. Experimental testing times were added between 4 and 20 hours.

Testing Times (Hours)	Number of Subjects	Number of Time Points
4 time-point designs		
0,2,24,168	10	4
0,4,24,168	10	4
0,8,24,168	10	4
0,12,24,168	10	4
0,16,24,168	10	4
5 time-point designs		
0,2,4,24,168	10,20,30	5
0,2,8,24,168	10,20,30	5
0,2,12,24,168	10,20,30	5
0,2,16,24,168	10,20,30	5
0,2,20,24,168	10,20,30	5
Alt. 5 time-point designs		
0,4,12,24,168	10,20,30	5
0,8,12,24,168	10,20,30	5
0,4,8,24,168	10,20,30	5
Extended time-point designs		
0,2,4,12,24,168	10	6
0,2,8,12,24,168	10	6
0,2,4,8,12,24,168	10	7
0,4,8,12,16,24,168	10	7
0,2,4,8,12,16,24,168	10	8
0,2,4,8,12,16,20,24,168	10	9

6.3 Simulation Results

6.3.1 Four Time Point Designs: Ten Subjects

In the 4-time point design, only the 2nd point differs from the original FOB experiment protocol. The 2nd time point varied from 2 to 16 hours, the results showed pronounced effect on the power of detecting the parameter K_{eo} (Table 5). When the second time point was chosen to be hours 8 or 12, the statistical power for the non-zero K_{eo} improved from 45 % to approximately 80%. However, for the parameters B and C the power of detecting a non-zero value remained low, ranging from 0.1% to 15.5%. The convergence rate in fitting the biphasic model decreased markedly as the second testing time moved past 12 hours. For example, the convergence was only about 72.9% at 16 hours, suggesting the data did not provide adequate information to even fit the model. Note that the shaded row indicates the design that appeared to be most beneficial for recovering the true model response.

Table 5. Summary of simulated design trials with 10 subjects per dose group and 4 testing times.

4 Point Designs	PARAMETER Power (% p<0.05) (average p value of t-statistic)			
Time Points	B	C	K_{eo}	Convergence (%)
0,2,24,168	15.5(0.329)	0.0(0.466)	45.5(0.118)	94.1
0,4,24,168	11.5(0.309)	0.0(0.392)	69.1(0.068)	93.1
0,8,24,168	5.7(0.259)	0.0(0.329)	81.6(0.057)	88.5
0,12,24,168	1.2(0.271)	0.0(0.341)	80.5(0.063)	82.6
0,16,24,168	0.1(0.331)	0.0(0.398)	75.7(0.091)	72.9

Cont'd next page

Table 5 Cont'd

1) 2nd testing time was selected at various times from 2-16 hours and effects on power and statistical significance examined.
2) Numbers in columns 2, 3, and 4 indicate the percent of time a p value <0.05 was observed while numbers in parentheses represent the average p_value based on 1000 replications.
3) The first row (design 0,2,24,168) is the true FOB design on which this simulation was based

Bias and MSE were also used to evaluate an efficient design for parameter estimation using the biphasic model. Bias represents the average difference in the location of the parameter estimate while MSE represents the imprecision of that estimate across the 1000 replications of each simulated experimental design. Consistent with the results of the power analysis, bias and MSE for the K_{eo} parameter were the smallest at 8 hours and therefore represented the best choice of testing times within this group of designs (Table 6) for this endpoint. However, the bias and MSE in parameters B and C appeared to be the smallest when the second testing time was at 16 hours. Convergence decreased markedly for designs with the hour 16 being a testing time, which may have affected the MSE for this design by eliminating some of the larger variations in the parameter estimates. The MSE for the hour 8 testing time was not markedly different from the hour 16 design for parameters B, C and K_e and had a higher convergence rate. This lends further support for the 0,8,24,168 design as the best choice among the 4 time point designs. In general, the MSE was so large for the B parameter that it should be no surprise that statistical

significance was not achieved for this parameter in any of these designs. In view of the results in Table 6, one can conclude that 4-point designs did not have the statistical power to reliably detect the appropriate biphasic model for this endpoint.

Table 6. Bias and MSE for each parameter of interest for the 4 time point designs with 10 subjects.

Design	Bias_B	Bias_C	Bias_K_e	Bias_K_{eo}
0,2,24,168	-1.59E+03	8.91E+00	6.05E-03	7.47E-02
0,4,24,168	-1.76E+04	9.81E+01	1.12E-02	2.88E-02
0,8,24,168	-1.33E+02	6.59E-01	1.12E-03	-1.89E-02
0,12,24,168	-1.29E+02	6.48E-01	-1.09E-02	-6.43E-02
0,16,24,168	-9.87E+01	5.29E-01	-2.80E-02	-1.10E-01
Design	MSE_B	MSE_C	MSE_K_e	MSE_K_{eo}
0,2,24,168	6.55E+08	2.25E+04	3.03E-03	8.81E-02
0,4,24,168	7.94E+10	2.50E+06	4.34E-03	3.39E-02
0,8,24,168	3.53E+05	8.69E+00	3.16E-03	2.10E-02
0,12,24,168	1.63E+06	3.76E+01	4.52E-03	2.35E-02
01624168	3.08E+05	8.52E+00	6.59E-03	3.39E-02

6.3.2 Five Time Point Designs: Ten Subjects

Due to the lack of statistical power in parameter estimation in the 4-time point designs, we examined the benefits of adding an additional testing time. We retained the 4-time points of the FOB design and allocated an additional testing time between 4 and 20 hours. In a second situation of 5-point designs we also removed the hour 2 time point and added 2 time points between 4 and 12 hours to

see if time points closer to the true TOPE would increase the statistical power.

The 5-point designs generally improved the power for all parameters of interest (Table 7). The average p value for K_{eo} reached a level below 0.05, and the power improved to above 80% for designs with the additional testing times between 4 and 12 hours. The improvement in power was less pronounced when the additional point was beyond 16 hours. For parameter C, the power remained at a low level of less than 10%. For the B parameter, the power reached a highest level of 28% when the added time point was at 16 hours, and stayed at a comparable level within the time range of 8 to 20 hours.

Table 7. Summary of simulated design trials on FOB parathion data with 10 subjects per dose group and 5 testing times.

5 Point Designs	PARAMETER % p<=0.05 (mean)			
Time Points	B	C	K_{eo}	Convergence (%)
0, 2, 4, 24, 168	21.2(0.220)	2.0(0.303)	80.9(0.030)	97.3
0, 2, 8, 24, 168	24.5(0.147)	9.8(0.213)	82.4(0.030)	99.0
0, 2, 12, 24, 168	24.5(0.147)	9.8(0.213)	82.0(0.033)	98.4
0 ,2, 16, 24, 168	27.7(0.158)	7.4(0.233)	73.5(0.047)	97.2
0, 2, 20, 24, 168	25.8(0.198)	2.6(0.292)	58.3(0.072)	96.4
Alternate 5 point designs				
0, 4,12, 24,168	22.0(0.160)	7.8(0.222)	87.4(0.033)	95.2
0, 8,12, 24,168	17.6(0.144)	1.6(0.204)	89.6(0.032)	91.3
0, 4, 8, 24,168	21.2(0.169)	5.4(0.232)	90.4(0.024)	96.4

1) The experimental testing time (3rd testing time) was adjusted by 4 hour intervals and effects on power examined.

2) Numbers in columns 2, 3, and 4 indicate the percent of time a p value <0.05 was observed while numbers in parenthesis represent the average p_value based on 1000 replications.
3) The true design on which this simulation was based included actual testing times of 0,2,24, and 168 hours.

Interestingly, dropping the hour 2 testing time and adding two testing times between 4-12 hours had little effect on the statistical power though in one case (0,4,8,24,168) a large bias was introduced (Table 8). The bias for B, C, and K_e was smallest with the design of 0,8,12,24,168 hours while the bias of K_{eo} was the smallest with the 0,4,8,24,168 design. The MSE was smallest with the 0,8,12,24,168 design for three of the four parameters of interest.

Table 8. Bias and MSE for each parameter of interest for the 10 subject, 5 time point design.

Design	Bias_B	Bias_C	Bias_K_e	Bias_K_{eo}
0,2,4,24,168	-3.50E+02	1.86E+00	1.00E-02	4.11E-02
0,2,8,24,168	-1.33E+02	6.95E-01	9.53E-03	4.80E-02
0,2,12,24,168	-1.33E+02	6.93E-01	8.68E-03	4.56E-02
0,2,16,24,168	-1.23E+02	6.41E-01	5.04E-03	5.52E-02
0,2,20,24,168	-1.17E+02	6.05E-01	3.89E-03	7.47E-02
0,4,12,24,168	-5.71E+02	3.11E+00	5.77E-03	1.96E-02
0,8,12,24,168	-6.99E+01	3.58E-01	-4.18E-04	-2.22E-02
0,4,8,24,168	-1.47E+08	8.49E+05	8.97E-03	1.56E-02
Design	MSE_B	MSE_C	MSE_K_e	MSE_K_{eo}
0,2,4,24,168	3.39E+07	1.06E+03	1.81E-03	3.43E-02
0,2,8,24,168	1.28E+06	3.70E+01	9.67E-04	5.17E-02
0,2,12,24,168	1.28E+06	3.70E+01	1.08E-03	5.22E-02
0,2,16,24,168	2.94E+06	8.05E+01	1.31E-03	5.20E-02
0,2,20,24,168	3.55E+05	9.73E+00	1.67E-03	7.02E-02
0,4,12,24,168	2.21E+08	6.78E+03	1.91E-03	3.16E-02
0,8,12,24,168	1.68E+05	4.05E+00	2.40E-03	1.91E-02
0,4,8,24,168	2.16E+19	7.22E+14	2.36E-03	2.11E-02

6.3.3 Designs with Six – Nine Time Points.

The results of the 5-time point designs begs the question "How many testing times will be enough to achieve required statistical power for the B and C parameters and acceptable bias and MSE?". In an attempt to answer this question, we increased the number of testing times from 5, to between 6 to 9 points, by adding additional points between 0 and 24 hours. As the number of testing times increased, the statistical power improved but even with 9 time points, the statistical power was still below 70% for B and 37% for C (Table 9). It is clear that as the number of time points increases, the power increases generally. Furthermore, fixing the number of time points yielded some data more informative than others.

Table 9. Summary of simulated experiments with six to nine points and 10 subjects per dose group.

Time Points (Extended time point designs)	PARAMETER % p<=0.05 (mean)			
	B	C	K_{eo}	Convergence (%)
0,2,4,12,24,168	33.5(0.113)	13.2(0.171)	95.4(0.010)	98.8
0,2,8,12,24,168	43.7(0.084)	16.1(0.133)	93.2(0.017)	98.8
0,4,8,12,24,168	34.8(0.101)	12.6(0.152)	93.5(0.019)	96.9
0,2,4,8,12,24,168	47.4(0.073)	19.7(0.117)	97.3(0.006)	99.2
0,4,8,12,16,24,168	49.0(0.070)	18.8(0.114)	93.9(0.014)	96.8
0,2,4,8,12,16,24,168	62.6(0.049)	31.0(0.086)	98.0(0.004)	99.2
0,2,4,8,12,16,20,24,168	68.3(0.043)	36.6(0.077)	98.0(0.005)	99.0

1) Two experimental testing times were added (3rd and 4th testing times) and adjusted by 4 hour intervals and effects on power examined.
2) Numbers in columns 2, 3, and 4 indicate the percent of time a p value <0.05 was observed while numbers in parenthesis represent the average p_value based on 1000 replications.
3) The true design on which this simulation was based included actual testing times of 0,2,24, and 168 hours.

The bias and MSE were generally reduced as the number of testing times increased for parameters B, C, and K_e, although the smallest bias in K_{eo} was attained with the 7 time point design of 0,4,8,12,24,168 (Table 10) and smallest MSE for K_{eo} was associated with the 8 time point design 0,2,4,8,12,16,24,168. Even with the 9 time point design, the coefficient of variation for parameter B was still 45% of the average.

Table 10. Bias and MSE for Simulated Experiments with 6 and 9 time points and 10 subjects

Design	Bias.B	Bias.C	Bias.K_e	Bias.K_{eo}
0,4,8,12,24,168	-6.00E+01	3.06E-01	6.42E-03	1.37E-02
0,4,8,12,16,24,168	-7.32E+01	3.96E-01	4.63E-03	1.15E-02
0,2,4,8,12,24,168	-8.85E+01	4.62E-01	9.12E-03	1.76E-02
0,2,4,8,12,16,24,168	-4.94E+01	2.58E-01	7.59E-03	1.77E-02
0,2,4,8,12,16,20,24, 168	-3.99E+01	2.10E-01	6.37E-03	2.14E-02
Design	MSE.B	MSE.C	MSE.K_e	MSE.K_{eo}
0,4,8,12,24,168	3.46E+04	8.02E-01	1.22E-03	2.24E-02
0,4,8,12,16,24,168	8.57E+05	2.74E+01	1.18E-03	1.71E-02
0,2,4,8,12,24,168	1.29E+06	3.77E+01	7.05E-04	1.69E-02
0,2,4,8,12,16,24,168	7.05E+04	1.97E+00	5.94E-04	1.44E-02
0,2,4,8,12,16,20,24, 168	1.09E+04	2.71E-01	5.82E-04	1.84E-02

6.3.4 Designs with Increasing Number of Subjects

While the simulations have clearly demonstrated that increasing the number of time points and selecting "informative" time points can generally improve the efficiency of an experiment, there remains a considerable amount of MSE, particularly associated with parameter B.

This suggests that while selection of time points is essential in soliciting information on the shape of dose-response, to reduce variation increasing the number of subjects may also be important. Therefore, a second simulation was performed using the 5-point experiments with 20 or 30 subjects per dose group.

6.3.5 Five time point designs with twenty subjects

Adding 10 additional subjects to each dose group resulted in pronounced improvement in statistical power for estimating the parameters. It is seen from Table 11 that the power for K_{eo} was consistently above 95% among all experiments, and the power for parameters B and C reached 77% and 44% respectively under the design of 0,8,12,24,168 hours.

Table 11. Summary of simulated design trials on FOB Parathion data with 20 subjects per dose group and 5 testing times.

Time Points 5 Point Design (20 subjects per dose group)	PARAMETER % p<0.05 (average p value)			
	B	C	K_{eo}	Convergence (%)
0,2,4,24,168	33.0 (0.104)	10.1(0.168)	98.4 (0.004)	99.4
0,2,8,24,168	63.8 (0.049)	24.2 (0.089)	98.6 (0.003)	99.8
0,2,12,24,168	70.3 (0.041)	29.2 (0.081)	96.9 (0.006)	99.5
0,2,16,24,168	61.5 (0.053)	25.3 (0.100)	96.8 (0.009)	99.6
0,2,20,24,168	47.7(0.079)	18.2(0.142)	93.7(0.016)	99.0
0,4,12,24,168	64.3(0.048)	28.3 (0.086)	98.2 (0.004)	98.6
0,8,12,24,168	76.9 (0.036)	43.7 (0.067)	96.4 (0.011)	97.0
0,4,8,24, 168	57.0(0.056)	25.5 (0.097)	99.4 (0.002)	98.7

Cont'd next page

Table 10 Cont'd

1) The experimental testing time (3rd testing time) was adjusted and effects on power examined.
2) Numbers in columns 2,3, and 4 indicate the percent of time a p value <0.05 was observed while numbers in parenthesis represent the average p_value based on 1000 replications.
3) The true design on which this simulation was based included actual testing times of 0,2,24, and 168 hours.

Under these 20 subject experiments, bias associated with the estimates of B, and C was rather invariant while bias in K_e was minimized with the 0, 2, 20, 24, 168 design and K_{eo} was minimized with the 0, 2, 8, 24, 168 design. The MSE for the 20 subject designs was generally reduced by at least an order of magnitude compared with the same designs of half the number of subjects. The smallest MSE was achieved with designs where the third time was located at either 8 or 12 hours (Table 12). The best statistical properties of the K_{eo} parameter again was achieved under a slightly different design with time points of 0, 8, 12, 24, 168. While increasing the number of subjects per dose group did not reduce bias, it dramatically reduced the uncertainty of the statistical quantities as measured by MSE. Because of the disparity of statistical power in estimating the parameters, we further investigated the case of 30 subjects per dose group.

Table 12. Bias and MSE for each parameter of interest for the 20 subject, 5 time point design.

Design	Bias.B	Bias.C	Bias.K_e	Bias.K_{eo}
0,2,4,24,168	-5.39E+01	2.74E-01	8.34E-03	2.03E-02
0,2,8,24,168	-3.96E+01	2.08E-01	7.67E-03	1.34E-02

Cont'd next page

Table 12 Cont'd

0,2,12,24,168	-3.20E+01	1.73E-01	5.46E-03	2.61E-02
0,2,16,24,168	-3.17E+01	1.70E-01	4.83E-03	3.66E-02
0,2,20,24,168	-3.17E+01	1.68E-01	2.96E-03	4.94E-02
0,4,12,24,168	-3.53E+01	1.90E-01	5.41E-03	9.84E-03
0,8,12,24,168	-3.35E+01	1.83E-01	3.29E-03	-2.88E-03
0,4,8,24,168	-4.29E+01	2.26E-01	7.11E-03	3.03E-03
Design	MSE.B	MSE.C	MSE.K_e	MSE.K_{eo}
0,2,4,24,168	1.61E+04	3.65E-01	6.46E-04	1.30E-02
0,2,8,24,168	6.43E+03	1.46E-01	3.81E-04	1.29E-02
0,2,12,24,168	5.18E+03	1.21E-01	3.82E-04	2.05E-02
0,2,16,24,168	6.44E+03	1.47E-01	4.22E-04	2.79E-02
0,2,20,24,168	9.26E+03	2.08E-01	5.88E-04	3.21E-02
0,4,12,24,168	5.68E+03	1.29E-01	6.07E-04	1.00E-02
0,8,12,24,168	8.01E+03	1.80E-01	9.13E-04	8.36E-03
0,4,8,24,168	8.10E+03	1.86E-01	6.52E-04	6.69E-03

6.3.6 Five-Time Point Designs: Thirty Subjects

A final simulation was conducted on the parathion data using 30 subjects per dose group. As expected, adding an additional 20 subjects to the original design resulted in greater convergence rates and nearly satisfactory statistical power for all parameters. Table 13 clearly shows that the design of 0, 8,12,24,168 yielded the largest power for parameters B and C. We further note that the power of C is particularly sensitive to design whereas B, K_e, and K_{eo} are less so.

Table 13. Summary of simulated design trials on FOB Parathion data with 30 subjects per dose group and 5 testing times.

Time Points 5 Point Design (30 subjects per dose group)	PARAMETER % p<0.05 (mean)			
	B	C	K_{eo}	Convergence (%)
0,2,4,24,168	50.5(0.057)	21.1(0.103)	100(<0.001)	100
0,2,8,24,168	93.0(0.016)	73.4(0.038)	99.8(<0.001)	99.9
0,2,12,24,168	94.6(0.014)	78.2(0.035)	99.9(0.001)	100

Cont'd next page

Table 13 Cont'd

0,2,16,24,168	89.4(0.021)	56.7(0.049)	99.8(0.002)	99.9
0,2,20,24,168	70.4(0.039)	35.3(0.081)	99.5(0.003)	99.9
0,4,12,24,168	92.7(0.016)	73.0(0.037)	97.8(0.002)	98.0
0,8,12,24,168	97.8(0.009)	92.0(0.023)	97.0(0.004)	98.5
0,4,8,24,168	88.5 (0.021)	67.3 (0.043)	99.6 (<0.001)	99.9

1) The experimental testing time (3rd testing time) was adjusted by 4 hour intervals and effects on power examined.
2) Numbers in columns 2, 3, and 4 indicate the percent of time a p value <0.05 was observed while numbers in parenthesis represent the average p_value based on 1000 replications.
3) The true design on which this simulation was based included actual testing times of 0,2,24, and 168 hours

We expect that adding additional subjects to the experimental design will reduce the MSE but that the relative performance of the designs will remain consistent with the results under 20 subjects. Indeed, as with the twenty-subject experiments, the 0,2,20,24,168 design with 30 subjects retained the smallest bias for parameters B, C, and K_e while the 0,8,12,24,168 design has the smallest bias for K_{eo} (Table 14). It is intriguing that, except for K_e, bias was reduced under the 30-subject designs, although the magnitude was small. A plausible explanation is the replication size (1000) of the simulation may not be sufficiently large to stabilize the variation of the parameter estimators.

The design which attained the smallest MSE for K_e was consistent with the 20-subject design (i.e. 0, 2, 12, 24, 168). However, the MSE for parameters B and C and K_{eo} was slightly different (though the same order of magnitude) from the results using 20 subjects. Specifically, the smallest MSE for these parameters occurred with the

0, 8, 12, 24, 168 design under 30 subjects. This discrepancy may be attributed to (1) reduction in bias and (2) non-convergence in simulations. Reduction in bias may change the expected relationship between MSEs' of the 20- and 30-subject designs. When a simulated experiment resulted in non-convergence in model fitting, the underlying extreme value was excluded in the computation of bias and MSE. Since the non-convergence rate was slightly higher under the 20-subject design, it is possible that the designs with the smallest observed MSE would no longer be the one if the convergence rate improved. Finally, there is a possibility that chance alone was responsible for the inconsistent MSE. If this is the reason, increasing the size of simulation replication would be helpful. Despite this artifact of inconsistency, it is clear that including testing times around the true TOPE was most beneficial in recovering the true dose-response profile.

Table 14. Bias and MSE for each parameter of interest for the 30 subject, 5 time point design.

Design	Bias.B	Bias.C	Bias.K_e	Bias.K_{eo}
0,2,4,24,168	-4.70E+01	2.41E-01	8.87E-03	6.82E-03
0,2,8,24,168	-3.21E+01	1.72E-01	6.84E-03	2.15E-03
0,2,12,24,168	-2.71E+01	1.48E-01	5.63E-03	8.42E-03
0,2,16,24,168	-2.55E+01	1.39E-01	4.77E-03	1.75E-02
0,2,20,24,168	-2.53E+01	1.37E-01	3.69E-03	2.99E-02
0,4,12,24,168	-2.80E+01	1.55E-01	3.50E-03	5.85E-03
0,8,12,24,168	-2.54E+01	1.43E-01	3.74E-03	-1.21E-03
0,4,8,24,168	-4.29E+01	2.26E-01	7.11E-03	3.03E-03
Design	MSE.B	MSE.C	MSE.K_e	MSE.K_{eo}
0,2,4,24,168	1.01E+04	2.30E-01	3.99E-04	4.37E-03
0,2,8,24,168	3.44E+03	7.98E-02	2.46E-04	4.03E-03

Cont'd next page

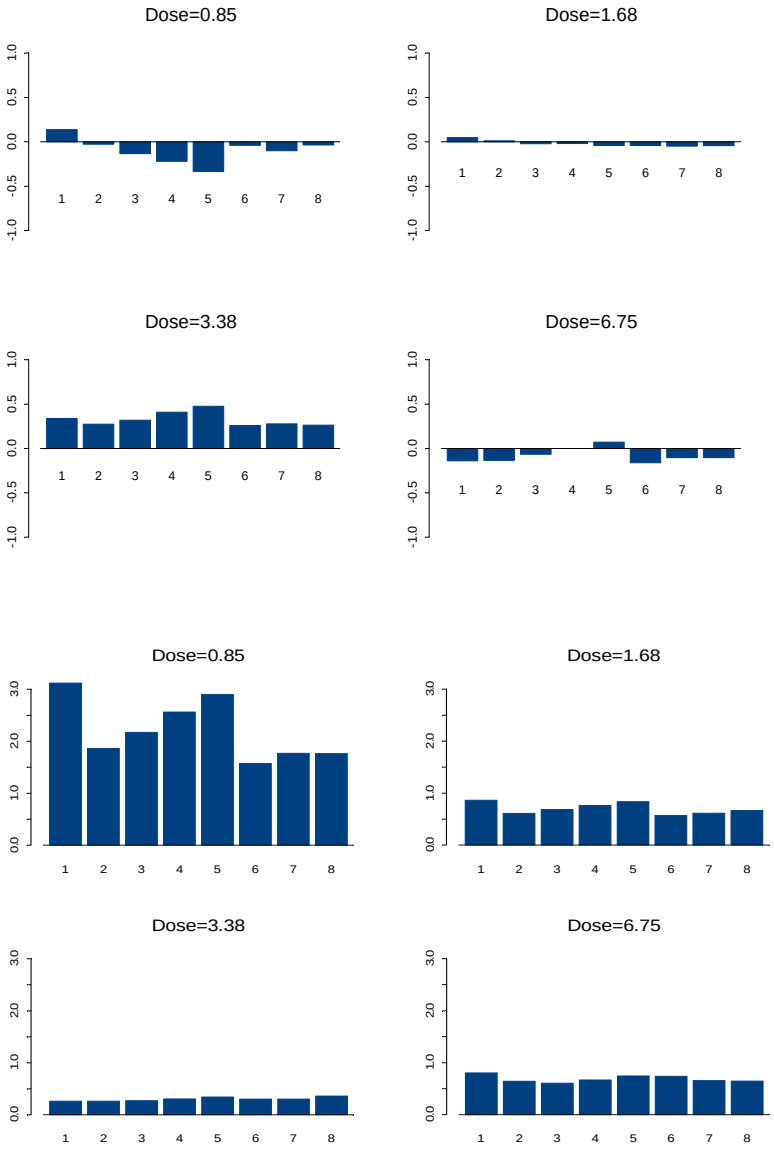
Table 14 Cont'd

0,2,12,24,168	3.33E+03	8.01E-02	2.00E-04	5.67E-03
0,2,16,24,168	3.44E+03	7.92E-02	2.57E-04	1.12E-02
0,2,20,24,168	5.43E+03	1.26E-01	3.01E-04	1.32E-02
0,4,12,24,168	3.52E+03	8.02E-02	6.18E-04	2.18E-02
0,8,12,24,168	2.21E+03	5.03E-02	4.97E-04	4.97E-03
0,4,8,24,168	4.73E+03	1.09E-01	2.97E-04	3.55E-03

6.3.7 Bias and MSE in Estimating Time of Peak Effect

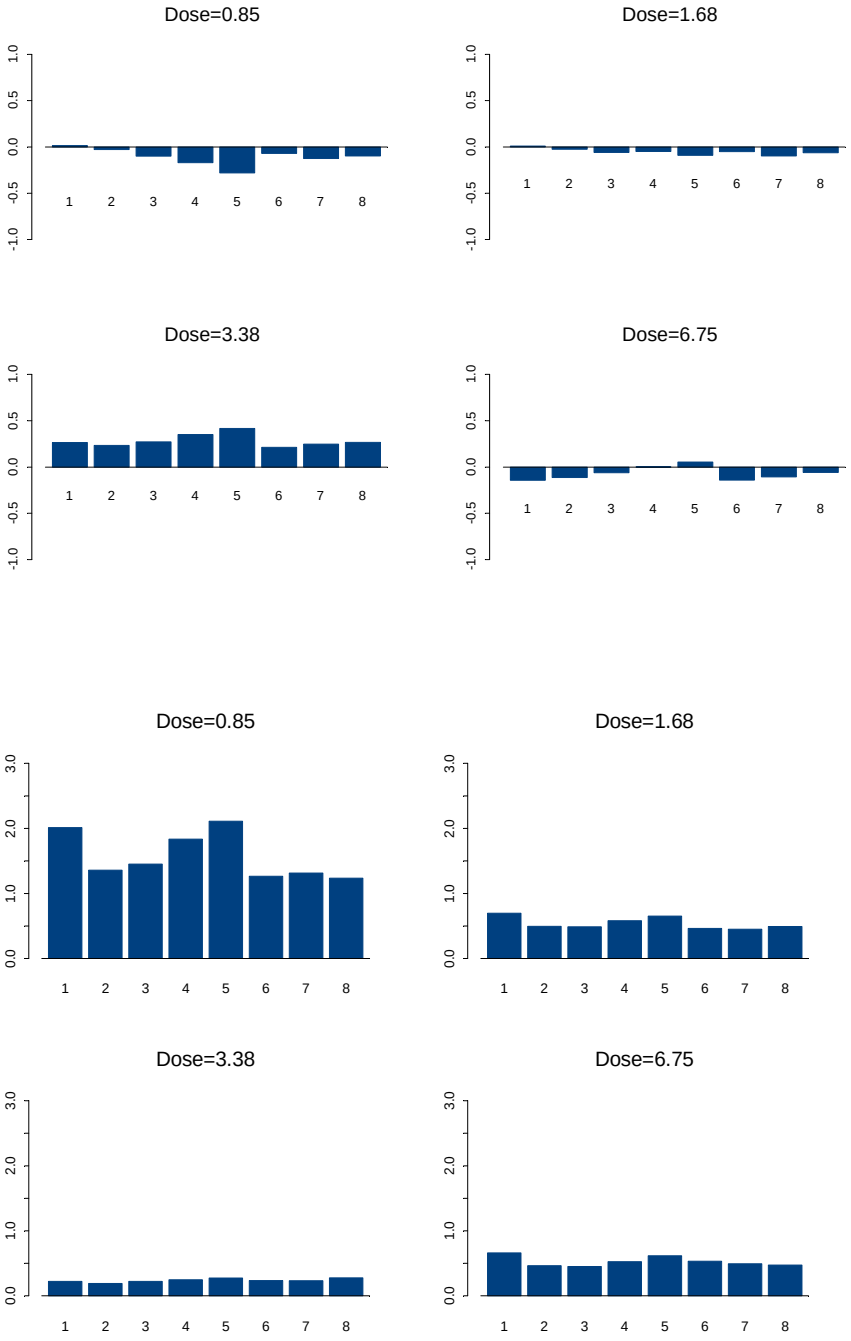
In addition to examining the Bias and MSE associated with the parameter estimates it is especially useful for the purposes of benchmark dose estimation to investigate efficient designs for estimating the TOPE. Since the TOPE is critical to benchmark dose estimation, Bias and MSE of the TOPE estimator are useful criteria for comparing designs. The 5 time point designs with 20 and 30 subjects per dose group were assessed with respect to identifying the design that recorded the smallest Bias and MSE for each dose group specific TOPE. For the both the 20 and 30 subject designs, the Bias was within one half hour of the true TOPE for each of the dose groups. The MSE appeared relatively invariant to design for all dose groups except for the lowest dose group (0.85mg/kg) which had the highest magnitude in MSE overall (Figures 13 and 14) and seemed to benefit most from designs where the testing times were closest to the true TOPE. This lowest dose group is especially important to consider because of its

relevance in establishing a Reference Dose used to define safe exposure limits.



Design Key: 1=0,2,4,24,168; 2=0,2,8,24,168; 3=0,2,12,24,168
4=0,2,16,24,168; 5=0,2,20,24,168; 6=0,4,8,24,168; 7=0,4,12,24,168;
8=0,8,12,24,168

Figure 13. Bias (Top) and Mean Square Error (Bottom) of the TOPE estimator under the 5-time point 20 subject designs.



Design Key: 1=0,2,4,24,168; 2=0,2,8,24,168; 3=0,2,12,24,168
4=0,2,16,24,168; 5=0,2,20,24,168; 6=0,4,8,24,168; 7=0,4,12,24,168;
8=0,8,12,24,168

Figure 14. Bias (Top) and Mean Square Error (Bottom) of the TOPE estimator under the 5-time point 30 subject designs.

Confidence intervals around the TOPE estimates were generated for each of the 5 time point, 20 and 30 subject replicate design trials to compare the ability of each design to recover the true TOPE (Table 15). As shown by the bias and MSE figures the variability in TOPE estimation was largest for the lowest dose group. As dose levels increased, the confidence intervals were within +/- one hour of the true TOPE (Table 15). The results of the TOPE analysis indicate that the biphasic model predicts the TOPE with validity (small bias) and that the reliability of the TOPE estimate (MSE) seems to be dependent on the number of subjects as well as the location of testing times, especially for the lowest dose group). Testing times located in proximity to the true TOPE for each dose group appeared to increase the reliability of the estimate for that dose group.

Table 15. Confidence intervals for dose group specific TOPE for each of the 5 time point, 20 subject designs. Confidence intervals were generated by calculating the 2.5% and 97.5% of the distribution of 1000 TOPE estimates for each design.

	Dose Group			
20 Subject Designs	0.85	1.69	3.38	6.75
0,2,4,24,168	12 (9-15)	9 (8-11)	7 (7-8)	7 (5-8)
0,2,8,24,168	12 (9-14)	9 (8-10)	7 (7-8)	7 (5-8)
0,2,12,24,168	12 (9-14)	9 (8-10)	7 (7-8)	7 (5-8)
0,2,16,24,168	12 (9-14)	9 (8-10)	7 (7-8)	7 (5-8)
0,2,20,24,168	12 (9-15)	9 (8-11)	7 (7-8)	7 (5-8)
0,4,8,24,168	12 (10-14)	9 (8-10)	7 (6-8)	7 (5-8)
0,4,12,24,168	12 (9-14)	9 (8-10)	7 (6-8)	7 (5-8)
0,8,12,24,168	12 (9-14)	9 (8-10)	7 (6-8)	7 (6-8)

Table 16. Confidence intervals for dose group specific TOPE for each of the 5 time point, 30 subject designs. Confidence intervals were generated by calculating the 2.5% and 97.5% of the distribution of 1000 TOPE estimates for each design.

	Dose Group			
30 Subject Designs	0.85	1.69	3.38	6.75
0,2,4,24,168	12 (10-15)	9 (8-11)	7 (7-8)	7 (5-8)
0,2,8,24,168	12 (10-14)	9 (8-10)	7 (7-8)	7 (6-8)
0,2,12,24,168	12 (9-14)	9 (8-10)	7 (7-8)	7 (6-8)
0,2,16,24,168	12 (9-14)	9 (8-10)	7 (7-8)	7 (6-8)
0,2,20,24,168	12 (9-14)	9 (8-10)	7 (7-8)	7 (6-8)
0,4,8,24,168	12 (9-14)	9 (8-10)	7 (7-8)	7 (6-8)
0,4,12,24,168	12 (10-14)	9 (8-10)	7 (7-8)	7 (6-8)
0,8,12,24,168	12 (10-14)	9 (8-10)	7 (6-8)	7 (6-8)

Chapter 7

Conclusions

In this thesis we have developed a biphasic toxico-diffusion model to describe dose-dependent time of peak effect (TOPE) in neurobehavioral toxicity screening studies. The model has been applied to a dataset from the IPCS collaborative study testing the neurotoxic potential of the chemical parathion using motor activity counts. The biphasic model predicted a decrease in motor activity counts as dose increased with a dose-accelerated TOPE: 11.5 hrs in the lowest dose group to 7.1 hrs in the highest dose group. The biphasic model was a significant improvement relative to the fit of the toxico-diffusion model according to the LRT as well as AIC and BIC. Due to the limited amount of data, however, estimates for parameters B and C were not statistically significant. Estimates of TOPE varied greatly between the two models with the toxico-diffusion model predicting a TOPE irrespective of dose at 26 hours while the biphasic model predicted a TOPE between 7-12 hours, decreasing with higher dose levels. Further, the magnitude of peak response (activity count at the TOPE) was dramatically different between the two models. Under

the toxico-diffusion model the minimum activity count was predicted to be 100 at the highest dose level, whereas under the biphasic model the peak response was predicted to be an activity count of 3 (the observed response in the data). This difference in magnitude could have dramatic impacts on benchmark dose estimation.

The simulation experiments identified the limitations of the existing FOB design protocol for eliciting physiologically relevant dose-response information and explored possible design improvements in order to generate adequate data for dose-response modeling. A specific focus of the simulation was to investigate the sensitivity of the experiments with respect to dose-dependent TOPE. Our results suggest that under the present experimental protocol of the FOB study, there is a substantial lack of statistical power when fitting a model such as the biphasic model. Only when the time spacing is adequate will information about the true shape in time be available to predict the dose-response relationship with validity. Does-dependent TOPE's cannot be verified if there are only limited time points in the experiment. In this regard, the existing protocol of FOB is unlikely to generate data that support dose-dependent TOPE. The simulation results further demonstrated that the design protocol can be altered to gain sensitivity and statistical power by adding additional time points and/or additional subjects to each dose group in the experimental

design. By adding a fifth experimental testing time point between hour 2 and 24, the sensitivity of the biphasic model improved with regard to estimating the dose-dependent time trajectory (e.g. TOPE) as witnessed by the increased model fitting convergence rate. By adding an additional time point, there are also more data points leading to an increase in statistical power for detecting the kinetic parameter K_{eo}, which determines the dose-dependent TOPE. Further increases in power can be achieved by adding additional subjects. It is debatable, however, whether it is practical to have at least 30 subjects per dose group in these screening tests. In suggesting design protocol improvements, consideration must be given to the practicality of any suggested design for implementation within the framework of the EPA testing guidelines for acute neurotoxicity.

An objective of the IPSC studies was to validate the study protocol across participating laboratories. As a result, several laboratories conducted experiments on the same chemical using identical methodologies. However, each individual laboratory identified the TOPE of the chemical through a pilot study and used that TOPE estimate as the second testing time for the FOB experiments. Since the definition of motor activity counts and the placement of the 2^{nd} testing time were left to the discretion of the individual laboratories, there are concerns about pooling data across laboratories for statistical

inference. Studies such as the IPCS would benefit from an experimental design which standardized the approach to TOPE estimation and motor activity counts in such a way that these data could be pooled across laboratories. The example below considers possible alternatives to the IPCS design that may effectively allow for more physiologically relevant information to be obtained without substantial investment of additional resources.

The IPCS collaborative studies rely on TOPE estimates from range finding studies using gate and arousal scores; however, there is evidence that different functional domains may exhibit different time courses following acute chemical exposure (Lammers and Kulig, 1997. Presuming that the TOPE for a given endpoint is measured without error, testing endpoints in more functional domains would yield a range of possible TOPE's for chemical exposure. This range of TOPE's for the given chemical could be used to guide time point selection. For example, even using only gait and arousal scores it is quite possible that the TOPE was different for these scores yet only one time point was chosen to represent the TOPE for the IPCS studies. As an alternative to the design above, a testing time associated with each TOPE (in this case 2) could be randomly assigned to participating laboratories. The data across laboratories could then be pooled resulting in a dataset where testing times for each TOPE estimate were

captured while each participating laboratory would only perform the experiments using 4 testing times. Mixed models such as those developed in this thesis allow for missing information from the individual time trajectories that would result from pooling these data. This flexibility a critical advantage of using mixed effects models for dose-response assessment compared to traditional methods such as analysis of variance. A scenario such as the example above could be expanded by considering a situation in which 5 metrics were used in the pilot studies to represent a neurobehavioral endpoint from each of 5 functional domains. A range of potential TOPE's could be established from these experiments. Each laboratory could be randomly assigned a testing time within the established range of TOPE's. Once the data were pooled, the resulting dataset would have numerous time points for analysis; a scenario more consistent with kinetic studies used to estimate physiological properties of the time course of a chemical in an organism after exposure. Toyinbo (2004) has shown that for functional domain composite scores, the precision of the TOPE estimate is somewhat insensitive to the actual time of TOPE but that the MSE appeared to be minimized when the time points were chosen prior to the actual time of peak effect. Our findings suggest that testing times bracketing the dose-dependent TOPE's minimized the imprecision of the TOPE estimator and increased the power of the

model. By adding even two time points bracketing the range of TOPE's from range finding studies, the experimental design would be capable of capturing more information for more neurobehavioral endpoints.

Another possible design protocol would be to choose distinct time points for different dose group in anticipation that the TOPE may dose-dependent. Since the range finding studies used a single dose for the TOPE estimate, if the chemical possesses dose-dependent TOPE, obviously a single testing time would capture the TOPE only for that dose. If information were available on potential dose-dependent TOPE for a given chemical under study, defining dose specific time points for the experiment would yield relevant dose-dependent information related to TOPE while maintaining the current sampling effort.

Adding additional subjects in each dose group would also improve estimates of the dose group specific average response which should increase the statistic power to detect significant response to chemical exposure, but only if the testing times are chosen appropriately would the benchmark dose estimation procedures be improved.

This thesis makes several simplifying assumptions in attempting to link mathematical models to pharmacologically relevant parameter estimates. Bioavailability of oral induced exposure to parathion is

assumed to be 100%. Further, absorption from the gastrointestinal tract is assumed to be in rapid equilibrium with concentration in the plasma compartment. From there it is the distributional delay associated with the hypothesized effects compartment that governs the dose-dependent TOPE observed in the data. There is some evidence that parathion and its toxic metabolite paraoxon is a chemical that induces non additive biochemical reactions in experimental subjects (INCHEM 2004). The biphasic toxico-diffusion imposed this dose-dependent TOPE but further study is necessary to validate this as a mechanistic product of parathion exposure.

The case of delayed TOPE was not illustrated in this thesis but a model was presented which may be useful in assessing dose-dependent TOPE arising from capacity limited kinetics. These situations have been observed in the literature (Dayneka et al. 1993, Jusko et al.,1995; Krzyzanski and Jusko 1997) though no examples of delayed effects with increasing dose were observed in the IPCS data.

The Nonlinear mixed effects models described in this thesis do not stand in isolation for assessment the neurotoxic potential associated with chemical exposure. Rather, the models serve to support a framework of different tools including hazard characterization, exposure assessment, risk characterization as well as

dose-response assessment utilized to build consensus on the potential risks associated with toxicity of the tested chemical. These tools are intended to support and not to replace the expert judgments of toxicologists who remain the key to understanding the implications of any inference of chemical exposure. This thesis has provided an additional tool that can be used in this regard and has illustrated its potential benefits when assessing neurotoxic potential of chemical exposure. It has further expanded the capability of risk assessment researchers to capture the diverse and sometimes complex aspects of dose-response relationships and contributed to the EPA's aims to use more physiologically relevant models to predict dose-response relationships over the time course of experimental studies.

List of References

Akaike, H (1973) Information theory and an extension of the maximum likelihood principle. In Proceedings of the Second International Symposium on Information Theory, B. N. Petrov and F. Csaki, eds. Akademiai Kiado, Budapest, pp. 267-281.

Anger, W.K. 1984. Neurobehavioral testing of chemicals: impact on recommended standards. Neurobehavioral Toxicology and Teratology. 6 147-153.

Anger, W.K. 1990 Worksite Behavioral Research: results, sensitive methods, test batteries and the transition from laboratory data to human health. Neurotoxicology 11:627-718

Dayneka, N.L., Garg,V., Jusko, W. J. 1993. Comparison of four basic models of indirect pharmacodynamic responses. Journal of Pharmacokinetics and Biopharmaceutics. 21(4):457-479

Cornish-Bowden, A. 1995. *Fundamentals of Enzyme Kinetics*. Portland Press, London.

Gabriellson, J. Weiner, D. 2000. *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications.* Swedish Pharmaceutical Press. Stockholm Sweden.

Holford, N.H., Sheiner, L.B. 1981. Understanding the dose effect relationship: clinical applications of pharmacokinetic-pharmacodynamic models. Clinical Pharmacokinetics. 6(6):429-453

INCHEM 1994. Guidance values for human exposure limits (EHC 170, 1994) : http://www.inchem.org/

Jusko, W. J. , Hui, C. K., Ebling, W. F. 1995. Convergence of direct and indirect pharmocodynamic response models. Journal of Pharmacokinetics and Biopharmaceutics. V 23 (1) 5-8

Kramer, R.E., Wellman,S.E. Rockhold,R.W. and Baker,R.C. 2002. Phamacokinetics of Methyl Parathion: A Comparison following Single Intravenous, Oral or Dermal Administration. Journal of Biomedical Science 9:311-320.

Krzyzanski, W. and Jusko, W.J. 1997. Mathematical formalism for the properties of four basic models of indirect pharmacodynamic responses. Journal of Pharmacokinetics and Biopharmaceuticals. 25(1):107-123

Kwon, Y. 2002 *Handbook of Essential Pharmacokinetics and Drug Metabolism for Industrial Scientists*. Kluwer Academic Publishers. New York.

Lammers, J.H. and Kulig, B.M. 1997. Multivariate time of peak effects assessment for use in selecting time of testing in acute neurotoxicity studies. Neurotoxicology 18(4):1079-1083.

Lindstrom, M., Bates, D. 1998 Newton Ralphson and EM algorithms for linear mixed effects models for repeated measures data. Journal of the American Statistical Association 4:1014-1022

Liu, T. (2000) Analyses of neurobehavioral screening data and risk assessment of neurotoxicity. Master's thesis, University of South Florida.

MacPhail, R.C., Peele, D.B., Crofton, KM. (1989) Motor activity and screening for neurotoxicity. Journal of the American College of Toxicology. 8:117-125.

McCarty, L.S. and Mackay, D. 1993. Enhancing ecotoxicological modeling and assessment: Body residues and modes of toxic action. Environmental Science and Technology 27: 1719-28.

Moser, V.C. Cheek, B.M., MacPhail, R.C. 1995. A multidisciplinary approach to toxicological screening III: Neurobehavioral toxicity. Journal of Toxicology and Environmental Health 45:173-210

Moser, V.C., Tilson, H.A., MacPhail, R.C., Becking, G.C., Cuomo, V., Frantik, E., Kulig, B., and Winneke, G. 1997a. The IPCS collaborative study on neurobehavioral screening methods: II. Protocol design and testing procedures. Neurotoxicology 18: 929-938

Moser, V.C., Becking, G.C., and Cuomo, V., Frantik, E., Kulig, B., MacPhail, R. C., Tilson, H.A., Winneke, G. , Brightwell,W.S., DeSalvia,M.A., Gill, M,W, Haggerty, G.C., Hornychova, M. Lammers, J. Larsson,J., McDaniel, K.L., Nelson, B.K. and Ostergaard,G. 1997b. The IPCS study on neurobehavioral screening methods: results of chemical testing. Neurotoxicology 18: 969-1056

Pinhiero,J.C. and Bates,D.M. 2000. Mixed Effects Models in S and S-plus. Springer-Verlag: New York.

Tilson, HA. (1990) Neurotoxicology in the 1990s. Neurotoxicology and Teratology 12:293-300.

Torda, T.A. Graham,G., Warwick,N.R. 1994. The Kinetics of Effect of Neuromuscular Blocking Drugs. The Keio Journal of Medicine. 43(1):27-30.

Toyinbo, P.A. On effective and efficient experimental design for neurobehavioral screening tests: The choice of a testing time for estimating the time of peak effects. Masters Thesis. University of South Florida.

U.S. Environmental Protection Agency (EPA) (1991) Pesticide assessment guidelines, subdivision F. Hazard evaluation: human and domestic animals. Addendum 10: Neurotoxicity, series 81, 82,and 83. Office of Prevention, Pesticides and Toxic Substances, Washington DC. EPA 540/09-91-123. Available from: NTIS, Springfield VA. PB91-154617

U.S. Environmental Protection Agency (1995) Reportable quantity adjustments; final rule. Federal Register 60(112):30926–30962.

U.S. Environmental Protection Agency (1996) Proposed guidelines for carcinogen risk assessment. Federal Register 61(79):17960-18011.

U.S. Environmental Protection Agency (1998) Guidelines for Neurotoxicity Risk Assessment Federal Register: 26926-26954

U.S. Environmental Protection Agency (2000) Benchmark dose technical guidance document, external review draft.

Vogel, J.S., Keating, G.A. Buchholz, B.A. 2002. Protein binding of isoflourophate in vivo after coexposure to multiple chemicals. Environmental Health Perspectives, 110 Suppl. 6: 1031-6. Abstract

Woodruff, S. 2001. The use of mixed effects dose-response models for neurotoxicity risk assessment. Masters Thesis. University of South Florida.

World Health Organization (WHO) Environmental Health Criteria 60: Principles and Methods for the assessment of Neurotoxicity Associated with Exposure to Chemicals. Geneva: World Health Organization, 1986.

Zhu, Y. (2001) Neurobehavioral Toxicity Risk Assessment: Dose-Time-Response Modeling and Benchmark Dose Estimation, US EPA Report

Zhu, Y. (2005a) Dose-time-response modeling of longitudinal measurements for neurotoxicity risk assessment. Environmetrics 16.(In press)

Zhu, Y., Wessel, M.R., Liu, T., and Moser, V.C. (2005b) Analyses of Neurobehavioral Screening Data I: Dose-Time-Response Modeling of Continuous Outcomes. Regulatory and Applied Toxicology. 41:240-255

Zhu, Y., Jia, Z., Wang, W., Gift, J., Moser, V.C., and B.J. Pierre-Louis (2005c), Data Analysis of Neurobehavioral Screening Data II: Benchmark Dose Estimation. Regulatory and Applied Toxicology (In press)