Analyses of Neurobehavioral Screening Data:
Dose-Time-Response Modeling of Continuous Outcomes

Yiliang Zhu†,1, Michael R. Wessel1, Tiebin Liu1,2, and Virginia C. Moser3

† Address correspondence to: Yiliang Zhu, Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, MDC56, Tampa FL 33612, USA, yzhu@hsc.usf.edu, (Tel) 813-974-6674, (fax) 813-974-4719
1 Department of Epidemiology and Biostatistics, University of South Florida, Tampa FL, USA
2 St Jude’s Children’s Hospital, Memphis, TN USA
3 National Health and Environmental Effects Research Laboratory, US EPA, Research Triangle Park, NC USA

Total word count: Abstract, 195; text, 5236; reference, 692

Disclaimer: The information in this document has been funded in part by the U.S. Environmental Protection Agency. It has been subjected to review by the National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.
Abstract. Neurotoxic effects are a non-cancer endpoint for health risk, and neurobehavioral screening tests can serve as a first tier investigation of neurotoxicity (US EPA, 1998). Analysis of neurobehavioral screening data such as those of the functional observational battery (FOB) traditionally relies on analysis of variance (ANOVA). ANOVA is designed to detect whether there are dose-effects, but does not model the underlying dose-response relationship and subsequent risk assessment fails to utilize the shape of the underlying dose-response. In contrast, dose-response modeling interpolates toxic effects between experimental points, and permits prediction of toxic effects within the experimental range. Additionally it is also a prerequisite for estimating a benchmark dose. This paper discusses dose-time-response modeling of longitudinal neurotoxicity data and illustrates the methods using three continuous FOB outcomes from an EPA study involving acute exposure to triethyltin (TET). Several mathematical functions are presented as candidate dose-time-response models. The use of random effects is discussed to characterize inter-subject variation. The results indicate that it is feasible to use simple mathematical functions to model empirical dose-time-response observed in existing longitudinal neurotoxicological data. Further research is needed on the types of design and data required to reliably approximate the true underlying dose-time-response.

Running Head: Dose-time-response modeling of continuous FOB data

Key words: Dose-time-response, functional observational battery, neurobehavioral toxicity, risk assessment, inter-subject variation, random effects
Introduction

Neurotoxicity caused by chemical exposure is among the ten leading causes of workplace disorders and its resultant harmful effects on human health are well documented (Anger, 1984; WHO, 1986; OTA, 1990; USEPA, 1998). Under the Toxic Substance Control Act (TSCA), the US Environmental Protection Agency (USEPA) has registered more than 65,000 chemical substances manufactured, imported, or processed in the United States, and this number is increasing every year (USEPA, 1996). An overwhelming majority of these chemicals have not been tested for neurotoxic potential (NRC, 1983). Under the mandate to regulate substances that have toxic potential to human health, the USEPA has developed guidelines that describe the principles, concepts and procedures for health risk assessment. The EPA guidelines for neurotoxicity risk assessment (USEPA, 1998) clearly identify the need for developing quantitative tools for assessing exposure-related neurotoxic effects in humans.

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). Neurotoxic effects can be observed at various levels of organization of the nervous system, including neurochemical, anatomical, physiological or behavioral. Neurobehavioral effects include adverse changes in somatic/autonomic, sensory, motor and/or cognitive function (USEPA, 1998); they can be reversible (the organism returns to pre-exposure condition) or irreversible (the organism is permanently and adversely changed by exposure). However, even reversible effects can signify underlying resultant damage to the organism (USEPA, 1998).
Neurotoxicity screening tests have been widely used as a first tier tool because they are simple, rapid, and economical, and represent the most fundamental level of investigation for neurotoxicity (WHO, 1986). A functional observational battery (FOB) in conjunction with an automated test of motor activity has been developed and used to examine a wide range of neurobiological functions, including sensory, motor, and autonomic function (see Moser et al., 1995; Moser et al., 1997a,b). Aiming to become a standardized neurotoxicity screening method, the FOB is non-invasive and rapidly generates behavioral data (WHO, 1986; USEPA, 1998) at baseline, during (for repeated exposure studies) and after the exposure period.

Analysis of the FOB as well as other neurobehavioral data has largely relied on the method of analysis of variance (ANOVA) with repeated measurements (Moser et al., 1995; Moser et al., 1997b). Recent advances in risk assessment methodology such as the benchmark dose method (Crump, 1984) require explicit dose-response modeling of toxicological data. With adequate data, dose-response modeling interpolates toxic effects between selected experimental points, hence predicting potential toxic effects as a continuum between experimental conditions that are otherwise un-measured. Risk assessment based on explicit dose-response modeling takes into consideration the variability of the data and the overall shape of the dose-response curve, and provides a more consistent basis for the calculation of reference doses (RfD) for regulation purposes (Gaylor et al., 1998).

The purpose of this paper is to introduce dose-time-response modeling for longitudinal neurotoxicity data, and to apply some mathematical models (Zhu, 2001; 2003; 2004) to continuous FOB measures. The paper focuses on statistical procedures for fitting
non-linear models, including model selection, residual diagnosis, random effects to adjust for inter-subject variation (in susceptibility and response), and the use of graphic tools. It illustrates these methods using a dataset from an EPA experiment with acute exposure to triethyltin (TET) (Moser et al., 1997b).

**Neurobehavioral Screening and a Functional Observation Battery**

The FOB test and resultant data considered in this paper utilized a single dose exposure at 4 levels in addition to a control. The testing took place 4 times, at baseline, an estimated time of peak effects (TOPE), one, and seven days after exposure. This sequence of test times differs somewhat from the latest EPA guideline (EPA, 1998), which specifies 7 and 14 days post exposure instead of 1 and 7 days as the last two testing times.

The FOB consists of 25-30 individual measures of multiple scales, including continuous, frequency/count, and nominal/ordinal (e.g. binary). This paper focuses on modeling the continuous outcome measures. Three continuous measures, hindlimb, forelimb grip strength, and landing foot splay, are used for illustration using the triethyltin bromide (TET) data.

TET is an organotin compound with widespread industrial uses as a stabilizer in certain plastic polymers used in the food industry. TET and trimethyltin are the most neurotoxic among the organotin compounds and are readily absorbed through the skin and gastrointestinal tract (Watanabe, 1980). These compounds target the central and peripheral nervous system (Cook et al., 1984a,b). The neurobehavioral toxicity of TET is characterized by impaired motor function resulting in hypoactivity, impaired grip strength and reduced amplitude of startle response (Squibb et al., 1980; Reiter et al., 1980). Because
of these well-documented neurotoxic effects, TET was selected as one of seven chemicals to serve as a positive-control in a multi-laboratory comparison of FOB data for use as a first tier screening mechanism for neurotoxicity assessment. Details of the test procedures are in Moser et al. (1997a).

**Dose-Time–Response Models**

Zhu (2004) consider a family of statistical models

\[ y = \eta(d,t) + e = A(t) + f(d,t) + e \]

to describe the behavioral trajectory of individual subjects. The observed response \( y \) at time \( t \) under the administered dose \( d \), is composed of two parts: the expected or average trajectory \( \eta(d,t) \) and the additive error \( e \). The expected behavioral trajectory \( \eta(d,t) \) further consists of two components: the first component \( A(t) \) is the natural behavioral trajectory in the absence of exposure (i.e. that of a control population); the second component \( f(d,t) \) represents dose-induced behavioral alterations that is additive to the reference trajectory \( A(t) \). This formulation imposes some constraints on the function \( f(d,t) \).

Because the baseline FOB test took place before dosing, there is no dose effect at time \( t=0 \), thus \( f(d,0) = 0 \) regardless of the dose level \( d \). Further, \( f(0,t) = 0 \) because there is no alteration to \( A(t) \) at time 0. Both \( A(t) \) and \( f(d,t) \) can be nonlinear functions of time or dose. Furthermore \( A(0) \) provides baseline behavior to serve as an internal control that is useful for adjusting for inter-subject variation. The functional form of \( A(t) \) and \( f(d,t) \) need to be fully specified, but the model coefficients involved are unknown and must be estimated.
from data through statistical procedures. Several mathematical options are available for \( A(t) \) and \( f(d,t) \).

Liu (2000) and Zhu (2001) report simple polynomial functions for dose-response modeling of FOB data. Higher order polynomial functions are flexible, but are extremely sensitive to local data variation. As a result, the interpolated dose-time-response between experimental points can be manifestation of the mathematical behavior of the polynomial function without data support or physiological justification. For example, a high order polynomial imposes multiple peaks or valleys on the dose-time-response, and is difficult to interpret. Therefore, polynomial functions beyond the second order in both dose and time are not recommended for practical purposes. The second order polynomial

\[
f'(d,t) = (B_1 d + B_2 d^2) t + (C_1 d + C_2 d^2) t^2
\]

is sufficient for most observed dose-time-response relationships in the acute exposure FOB data (Zhu, 2001). In this model there is no intercept because it is absorbed into the controls’ behavioral trajectory \( A(t) \). The term \( B_1 d + B_2 d^2 \) is the dose-dependent rate of change; and the term \( C_1 d + C_2 d^2 \) is the dose-dependent acceleration of change, or slope for time-squared. In a specific application statistical criteria can be used to determine if a reduced version of this function is adequate. For example, the rate of change and acceleration may be proportional to dose only, so squared dose terms are unnecessary, i.e. \( B_2 = 0 \) and \( C_2 = 0 \). The value of these coefficients is unknown and must be estimated statistically. While a polynomial model may fit the data adequately, its coefficients are mathematical quantities that describe the shape of the dose-time-response, and biological interpretation is often unavailable.

$$\eta(d, t) = A(t) + \frac{(B_0 + B_1 d)t}{1 + \exp(C_1 d)t^{\gamma_0}}$$

that resembles an extension of a diffusion process (Murray, 1993, Chapter 15) as well as the Michaelis-Menten enzyme kinetics equation. The case $D_0=1$ enables interpretation similar to the Michaelis-Menten equation. Specifically, the term $-(B_0 + B_1 d)\exp(-C_1 d)$ is the dose-dependent asymptote in time and $\exp(-C_1 d)$ is the dose-dependent half-life time. As such it prescribes a trajectory of persistent dose-effects, i.e. dose effects sustain at a constant level and do not show any recovery from the dose effects within the experimental time course; The case $D_0>1$ dictates that the term $\frac{(B_0 + B_1 d)t}{1 + \exp(C_1 d)t^{\gamma_0}}$ will diminish to zero in time regardless of the dose level, hence implying reversible or transient dose-effects. The case $D_0<1$ implies the dose-effects increase monotonically over time, a mathematical possibility but an unlikely case practically. With $d=0$, the function becomes

$$\eta(0, t) = A(t) + \frac{B_0 t}{1 + t^{\gamma_0}}$$

and describes the controls’ behavioral trajectory.

Zhu (2004) considered a general family of functions

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

that extends the Michaelis-Menten equation by incorporating dose- and time-dependent coefficients, $B*conc(d, t)$ and $C*conc(d, t)$. Whereas $conc(d, t)$ can be a general function of dose and time to be determined in specific applications, it is preferable to utilize
concentration at time \( t \) with administered dose \( d \), when feasible. This helps incorporate toxicokinetics in dose-response modeling. At the present stage, however, this family of functions remains mathematical and mechanistic interpretation is generally unavailable unless there are specific evidences supporting the use of particular toxicokinetics. For example, the one-compartment kinetics

\[
conc(d, t) = \frac{d}{V} \exp(-K_v t)
\]

is relevant only when it concerns concentration in blood circulation system under intravascular administration of dose. In neurotoxicity studies if exposure affects other organs (e.g. brain) of the body and lead to neurobehavioral changes, the one-compartment kinetics does not apply.

The coefficient \( C \) must be such that \( 1 + C * conc(d, t) * t > 0 \) to ensure the model is defined. During the experimental time-course, \( f(d, t) \) starts from zero, reaches a maximum effect value at the time of peak effect (TOPE), and returns back to zero (baseline) when \( conc(d, t) \) reduces to zero; it may fail to return to the baseline level when \( conc(d, t) \) remains greater than zero, prescribing persistent dose effects. Note that the time and/or magnitude of peak response generally depend on \( conc(d, t) \), hence the administered dose level.

Several special cases of this family of functions are employed in this paper. Letting \( conc(d, t) = d \exp(-K_v t) \), for example, yields

\[
f(t,d) = \frac{B d \exp(-K_v t)}{1 + C t d \exp(-K_v t)}.
\]

For convenience this function is called toxico-diffusion model (Zhu, 2004). The initial rate (first derivative with respect to \( t \)) of behavioral alteration (at time=0) is \( B d \), proportional to
dose level. Setting the coefficient $C=0$ give rise to the linear exponential function

$$f(t, d) = Btd \exp(-Kc_t).$$

Linearization of the toxico-diffusion function with respect to $Cdt \exp(-Kc_t)$ via a first order Taylor series expansion leads to the complementary-exponential function

$$f(t, d) = \exp(-Kc_t) \{1 - Ctd \exp(-Kc_t)\}$$

that is similar to the mutagenicity model of Myers et al. (1981). When $C > 0$, the term $1 - Cdt \exp(-Kc_t)$ represents a deterrent to dose effects; when $C < 0$, the term is an acceleration factor of dose effects. All these model coefficients are unknown in value and must be estimated from the data.

The toxico-diffusion function and its variations described above all start at zero initially ($t=0$), reaches a maximum (or minimum) at the time, $t = 1/Kc$, and then return to zero eventually (for sufficiently large $t$). Therefore they are appropriate for describing the dose-response trend with such a time-transient pattern. A limitation to a certain degree, these functions impose a common TOPE at $t = 1/Kc$ irrespective of administered dose level.

On the other hand, neurobehavioral screening experiments typically utilize just one time point between baseline and day 1 or day 7 (Moser et al. 1997a; USEPA, 1998), providing little information on how TOPE may be affected, if at all, by administered dose. Additional time points seem necessary in order to allow for dose-dependent TOPE.

**Materials and Methods**

*Data from the TET Experiment*
In the TET experiment (Moser et al., 1997a, b) rats (n=10/dose) were administered TET at dose levels of 0.75, 1.5, 3.0 and 6.0 mg/kg intraperitoneally, in addition to a vehicle control. Animals were tested using a FOB protocol at baseline (just before dosing), 2H (estimated TOPE), 24H, and 168H (7 days) after exposure. The TOPE was estimated through a pilot experiment using a single dose based on two endpoints, gait score and arousal (Moser et al., 1997a). Dose levels were also determined through the pilot study.

**Inter-subject Variation and Random Effects**

The dose-time-response modeling takes into consideration two major sources of data variation: inter- and intra-subject variation. The intra-subject variation is the error term $e$ in the model. The inter-subject variation can be attributed to biological difference among subjects and cannot be characterized by available prognostic or experimental factors. It is attractive to characterize the biological variation using random effects and separate it from the error. Characterizing biological variation is not only of interest in its own right but may also help improve statistical power. A common mechanism of random effects is to add a random coefficient $\alpha_i$ for each subject to selected model coefficients such as an intercept or slope. Models with random-effects are called mixed-effects models. Each random coefficient $\alpha_i$ quantifies the individual deviation from the group average response. Although in principle random effects can be attached to any model coefficient, the variation in the data may limit the choice. Therefore, the choice of random effects must reflect biological feasibility, statistical significance, and the actual amount of data variation. Only those of moderate to large size relative to the error contribute statistically to the model.
Model Fitting and Initial Parameter Values

Model fitting of mixed-effects models is more involved because of the latent nature of the random effects, and requires the use of numerical algorithms (e.g. the EM-algorithm, Newton-Raphson, or some hybrid combinations (Lindstrom and Bates, 1988)) to maximize the likelihood function. This is achieved through estimating the model coefficients at some tentatively fixed value of random effects and then update the random effects with the estimated model parameters. This process iterates until the estimates reach a converging point. Commercial software packages such as SAS (SAS Institute, Cary NC) and S-Plus (MathSoft, Seattle, WA) include routines for fitting mixed-effects models. The choice of initial values of the model coefficients to start these algorithms is crucial to achieving convergence of the iterative estimation process. A number of strategies can be employed to determine appropriate initial values (Zhu, 2003). Exploratory analysis of the raw data helps reveal the characteristics of the dose-time-response with respect to baseline level, direction of dose and time trend, slope, size, and time of peak effect, etc. These characteristics in turn provide important information about the initial values. Initial values are specific to the model and data at hand. They can be also obtained through solving a set of equations based on observed data, or fitting the model without random effects (random coefficients set to zero). Arriving at good initial values of nonlinear model parameters is more relevant than that of linear parameters. Pinheiro and Bates (2000) provide a technical discussion on initial values and model fitting.

Non-Constant Variances
Biological data tend to show non-constant variation in proportion to the changing level of mean response as a result of exposure. One can impose, thus, a distinct standard deviation for each dose level. For the FOB data used here, dose-specific standard deviation errors are estimated as a proportion of that for the controls, and the proportionality is reported. Likelihood ratio tests can be used to verify the significance of non-constant variances. The intra-subject variation may also be manifest in the form of serial correlation among data of the same subject. Various structures of serial correlation may be considered in analysis of longitudinal data (Pinhero and Bates, 2000, Chapter 8). In the present examples of FOB data, however, the few time points do not provide sufficient information to ascertain a specific correlation structure.

**Model Selection and Diagnostics**

Given a number of mathematical candidates, model selection is largely based on statistical criteria, including graphical comparison, likelihood ratio test and its generalizations (e.g. Akaike information criterion, AIC, and Bayes information criterion, BIC), and residual plots. These tools were used to examine distinct aspects of model fitting and to select a suitable model among a number of candidates. Because of the proximity between t=0 and t=2H compared to t=168H, a log-scale of time was used in all graphic displays to better reveal behavioral changes between 0 and 2H. Graphic tools were used to provide visual comparison between the observed dose-response and predicted responses from the mathematical model fitted to the data. Residuals, i.e., the difference between the predicted and observed response, were plotted to check the assumptions such as normality on data distribution, non-constant variances, or non-randomness of the errors. The
likelihood ratio test was used to compare two models of which one is a special case (sub-model) of the other. Related information criteria (AIC and BIC) were applied in situations where two comparison models follow the same family of distributions but one is not a sub-model of the other (Akaike, 1973). The information criteria penalize the inclusion of excessive model coefficients when their contribution to model is insignificant, and are coherent with the parsimony principle. The likelihood ratio tests and information criteria may not be valid for comparing two models, one with and another without random effects but otherwise identical because such tests may violate certain statistical conditions (Verbeke and Molenberghs, 2003).

Results

**TET Data Summary**

Table 1 presents a summary (sample mean and standard deviation by dose level and time point) for hindlimb grip strength, forelimb grip strength, and landing foot splay. Sample size was 10 per dose, except that one rat in the highest dose group died before 168H and an additional rat had missing data on forelimb and landing foot splay at 168H. Figures 1 and 2 (top row) and Figure 3a are spaghetti plots that show the observed trajectories of individual rats grouped by dose. In general, forelimb grip strength and hindlimb grip strength decreased after exposure, and larger and longer decreases occurred at higher dose levels. Hindlimb grip strength decreased among rats in the three highest dose groups, reaching a lowest level somewhere between 2H and 24H. By 168H it had recovered to a level near the baseline, except for the highest dose group in which recovery was about 55% of the baseline level on the average (Figure 1, top row). Forelimb grip
strength appeared to drop only among rats at the highest dose level and there was only limited recovery from the lowest level (at 24H) to baseline line level by 7 days (Figure 2, top row). Increased landing foot splay was seen mostly in the two highest dose groups, peaking immediately after exposure at around 2H, and recovering completely to the baseline level by 24H (Figure 3a). The dose-induced increase of landing foot splay was superimposed on its natural decrease among the controls. Repeated measurement ANOVA indicated significant overall dose effects on (1) hindlimb grip strength at all time points after exposure; (2) forelimb grip strength at all time points after exposure; and (3) landing foot splay at 2H after exposure (Moser et al., 1997b). Decreased hindlimb and forelimb grip strength has been consistently reported (Moser et al., 1997b; Squibb et al., 1980) while the effects of TET on landing foot splay are less reliable (Moser et al., 1997b; Reiter et al., 1980). The distinct trends of these three outcomes invite investigation of different functions for dose-response modeling.

*Hindlimb Grip Strength.*

The toxico-diffusion model was selected to characterize incomplete recovery of dose-time-response at higher dose levels. In comparing the toxico-diffusion model with its simplified version, the linear-exponential model, the likelihood ratio test gave a Chi-squared statistic value of 52.7 (p-value < 0.0001), suggesting that the denominator of the toxico-diffusion function (i.e. the coefficient C) makes significant contribution to the overall model fitting. Similarly, the likelihood ratio test favors the complementary-exponential model over the linear-exponential model (Chi-squared= 20.76, p-value<0.001). Although the likelihood ratio test is not applicable to comparing the toxico-diffusion and
complementary-exponential models, the information criteria (−120.50 vs. −88.55 for AIC and −100.74 vs. −68.79 for BIC) as well as the residual plot favored the toxico-diffusion model. The fitted toxico-diffusion model is displayed Figure 1 (bottom row), where rats are grouped in separate panels, one for each dose group. The fitted model stipulates a constant level of hindlimb grip strength for each control and implies reduced grip strength (B= -0.214, p=0.036, Table 2) as well as a reduced ability to recover (C=0.593, p=0.058) as dose level increases. It further reflects dose-related peak effects. Although the lowest hindlimb grip strength was observed at 24H, the parameter $K_e= 0.034$ suggests an estimated TOPE of 29.15 hours after dosing. While there were no data available at 29H to support the model-interpolated TOPE, estimation of TOPE is generally robust to the choice of the 2nd testing time when it deviates from the true TOPE so long as the mathematical model is a good approximation to the true model (Toyinbo, 2004). Thus, the critical issue is whether the experimental data are adequate and sufficient to uncover the true underlying dose-response. The impact of the random intercepts (baselines) can be seen in Figure 4 through the contrast between the individual (adjusted with random effects; dotted lines) and the group average trajectories (solid line). Overall the model describes the average observed dose-response. The predicted trajectory for a number of rats deviated visibly from the group average, suggesting a considerable degree of the inter-subject variation. The random-effects standard error is 0.089, about 60% of the residual standard error and 10% of the mean hindlimb grip strength (Table 2). Although the inter-subject variation in dose-response can be in principle characterized by additional random coefficient such as slope (coefficient B), statistical support for including more than one random coefficient was lacking. The use of random coefficient for A resulted in a model fit more satisfactory than B.
Residuals were plotted against fitted hindlimb grip strength in Figure 5 where each dose group is in a separate panel. There are no apparent signs of non-randomness. The range of predicted hindlimb grip strength is narrower for the controls than for the dose groups, echoing the exposure effects of possibly increasing variation. The residual plot also indicates a somewhat constant variance across dose groups, and the use of non-constant variances did not improve model fitting as judged by the likelihood ratio test (Chi-squared = 5.44, p-value=0.25). Inclusion of a serial correlation for errors of the same rat did not improve model fitting either (p-value = 0.78).

**Forelimb Grip Strength**

In fitting forelimb grip strength data the toxico-diffusion, complementary-exponential, and linear-exponential models resulted in similar values in maximum likelihood and information criteria, suggesting limited improvement of the toxico-diffusion model or complementary-exponential model over linear-exponential. However, graphical comparisons (e.g. residual plots) suggested a slightly better fit of the complementary–exponential model. The results in Table 3 show that using dose-specific variance improved the model fitting compared with a constant variance (p=0.048), with the standard deviation of the highest dose group being 1.43 times the control. The final model (Table 3) strongly suggested that dosing resulted in reduced forelimb grip strength \((B=-0.0073, \ p < 0.001)\). Although the coefficient \(C\) was only marginally significant (estimate \(C=0.0046, \ p = 0.06\)), it implies a possible lagging effect of TET on forelimb grip strength as compared with hindlimb grip strength. The estimated TOPE of 52.4 H \((K_e= 0.0191)\) suggests the same. Random intercepts were most effective in adjusting for intra-subject variation, and adding
additional random coefficients such as random slopes resulted nearly perfectly correlation between the two sets of random coefficients. Although significant, the standard deviation of the random intercepts is was only about 4% of the mean forelimb grip strength and 44% of the residual standard error. The estimated forelimb grip strength is displayed in a multi-panel spaghetti plot in Figure 2 (bottom row). The model predicts the TOPE at 52.4H whereas the observed peak effects were at 24H. Figure 6 provides a contrast between the fitted trajectory for each rat and the dose-group average in conjunction with the observed data points. The random effects account for only a small portion of the inter-subject variation as the trajectories for individual rats are clustered together around the group average. While the fitted model depicts clearly the average trend, it does not predict the individual trajectory indicated by the raw data of several rats. Using random effects on additional coefficients (i.e. B or C) may capture further inter-subject variation, but such efforts were not supported by the existing data. The residuals plot (Figure 7, 1st panel) shows a narrow range of predicted baseline levels even in the presence of random intercepts; it also reveals a larger variation of residuals in the highest dose group (Figure 7, 5th panel) under a distinct standard deviation for each dose group.

**Landing Foot Splay**

The average value of the controls decreased steadily until reaching a constant level after 24H (Table 1). The results in Table 1 also shows that at three highest dose levels, the average splay increased markedly at 2H after exposure, returned approximately to the control level by 24H, and stayed at a level comparable to that of the control. The trend clearly suggests dose effects as well as a complete recovery by day one. The spaghetti plot
of raw data (Figure 3a) shows a similar pattern that is discernable despite individual variation and a few exceptional rats (e.g. two rats in the highest dose level had elevated splay at 168H). Given a complete recovery, any data (e.g. 168H) after the recovery contributed no extra information about the dose-response. From a modeling viewpoint, it is difficult to devise a simple function to describe the dose-response shape of rapid-increase, recovery, and followed by a nearly flat level afterwards; from a risk assessment viewpoint, the data after the complete recovery provide no information on risk, magnitude, peak time, and duration. To circumvent non-convergence problem in fitting each model discussed here, therefore, data at 168H were removed from dose-response modeling. Still none of the toxico-diffusion, complementary-exponential, and linear exponential models adequately predicted the observed response. For example, the complementary-exponential model predicted a drop in splay between 2H and 24H to below the baseline level before rising back to the baseline again.

This led to the use of polynomial functions. The polynomial model involved a linear function $A_0 + A_1t$ to describe the controls’ changes in landing foot splay without exposure, and a quadratic time function $f(d,t) = B_1 dt + C_1 dt^2$ for dose-altered splay changes. The linear slope ($B_1$) of time characterizes increases in splay among rats in dosed groups and the quadratic slope ($C_1$) of time characterizes return to baseline level by 24H. These coefficients are proportional to dose to reflect the dose-effects. The model also includes random intercepts at the baseline to accommodate inter-subject variation. Table 4 is a summary of model fitting, and shows highly significant dose-effects ($p$-value < 0.0001 for both $B_1$ and $C_1$) and signifies the dose-response in 0-24H. The fitted model (Figure 3b) prescribes to dose-elevated splay among dosed rats compared with a steady decrease of the
controls. Standard deviation (12.38) of the random intercepts is 16% of the mean foot splay (76.32), but slightly larger than the residual standard deviation (11.47). Random effects of this size capture a substantial amount of inter-subject variation as can be seen from Figure 8 where individual trajectory of landing foot splay are shown in contrast to the group average as well as observed values. Incorporation of non-constant variance or serial correlation for errors did not improve model fitting. The fitted polynomial model dictates that TOPE=(-0.44+2.51d)/(0.21d), delayed TOPE with increasing dose. At the highest experimental dose (6 mg/kg), for example, the model yields TOPE=11.6H.

**Conclusions and Discussions**

We have presented several mathematical functions and discussed dose-time-response modeling for longitudinal neurotoxicological data of continuous scale, focusing on statistical process and potential issues. Although the examples utilize datasets from a single published experiment, the results have demonstrated the feasibility of explicit modeling of time-course dose-response data from the FOB tests. These and other applications to the FOB data (Zhu, 2001; 2003) lend support to the use of dose-response modeling, when data are adequate, as an alternative and supplement to traditional analysis of ANOVA. While ANOVA is effective to detect dose-induced changes or time-related variation in outcomes, it does not reveal or utilize the magnitude and shape of the underlying dose-response along the time course. Dose-response modeling, in contrast, attempts to overcome this limitation and is the first step to the deployment of the benchmark dose method for risk assessment.
The models discussed in this paper are mathematical in nature, and therefore do not have general physiological or toxicological interpretation. Kinetic studies have shown that TET is rapidly distributed to tissues, reaching an asymptote by about one hour with concentration level staying fairly constant throughout 24 hours (Cook et al., 1984a,b; Rose and Albridge, 1968). The dose-response models in the present paper appear to agree with these earlier studies, but the rapid recovery of landing foot splay and the lagging TOPE of forelimb grip strength seemed to reflect complex neurotoxicological mechanisms.

In selecting among these mathematical models, a primary consideration is that the mathematical function must match and accommodate the observed dose-response pattern. Statistical considerations are to ensure the selected model fit the data well with respect to the validity of statistical assumptions and a balance between simplicity and flexibility. Likelihood ratio test as well as the information criteria is useful for model selection within a hierarchical structure. They guide, for example, the inclusion or exclusion of additional model coefficients. Random effects can account for biological variation in response among subjects and also isolate it from the error variation. Since it is plausible that individual rats may exhibit distinct dose-response trajectories, it is appealing to include random effects on multiple model coefficients when data are adequate. In the examples of this paper, the amount of FOB data was insufficient to support the use of multiple sources of random effects; only the most effective one (e.g. baseline or intercept) was selected. Statistical tests for selecting random effects are technically involved and likelihood ratio tests are not always applicable (Verbeke and Molenberghs, 2003). Practical consideration can be directed to the size of variation of the random effects (i.e. standard deviation) relative to the mean response and residual variation. For instance, the size being less than 10% of the
mean or error variation implies limited effectiveness of using random effects. Estimation of true TOPEs is crucial not only because a TOPE is an important characteristic of the dose-response, but also it is associated with when maximum risk occurs. With only a few testing time points and only one targeting the TOPE, it is difficult to differentiate an observed TOPE from a true TOPE. In the absence of kinetic and toxico-dynamic information, it is also difficult to ascertain a model-interpolated TOPE between experimental time points. However, statistical design theory as well as a recent simulation study (Toyinbo, 2004) indicate that model-interpolated TOPE is often within a small neighborhood of the true one even under a selection of flexible testing times if a true model can be assumed. TOPEs can be dose-dependent, as allowed by the polynomial and diffusion models, but are verifiable only when observed data indicate so. The toxico-diffusion model and its variation impose a single TOPE and are empirically useful in the absence of dose-varying TOPE in the data.

It remains uncertain if the FOB data generated under the existing protocols (Moser et al. 1997a, USEPA, 1998) will ensure that the fitted model to be a reasonable approximation to the true underlying dose-response. Statistical theory suggests that the minimum number of experimental points (e.g. time points) must be equal to or greater than the number of model coefficients if the underlying model is known. When the underlying model is unknown, an even greater number of experimental points would be required. It is of interest to see how experiments should be designed to differentiate among several candidate models.

Upon adequately fitting a dose-response model, benchmark doses can be estimated. With longitudinal data, benchmark doses must reflect changing risk over the time course.
Benchmark dose computation for time-course data will be reported separately (Zhu et al. 2004).

**Acknowledgments** This work was supported in part by a National Science Foundation (DMS9978370) to Zhu and through an EPA Internship to Wessel. The authors thank Wei Wang, Zhihong Jia, Sanjay Munirredy, and Peter Toyinbo of the Health Risk Assessment Methodology Group at University of South Florida for helpful discussions. The results reported in this paper were partially presented at the *7th Biennial CDC/ATSDR Symposium on Statistical Methods*, January 1999, Atlanta, GA and the *International Conference International Conference on Statistical Challenges in Environmental Health Problems*, September 2001, Fukuoka, Japan.
References


Moser, V.C., Becking, G.C., 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., Hornyhova,
IPCS study on neurobehavioral screening methods: results of chemical testing.
*Neurotoxicology* 18, 969-1056.


the Process. Prepared by: Committee on the Institutional Means for Assessment of Risks to
Public Health, Commission on Life Sciences. Washington, DC.

Office of Technology Assessment (OTA), 1990. Neurotoxicity: identifying and controlling poisons
of the nervous system. U.S. Congress, Office of Technology Assessment (OTA-BA-436).

Verlag.

exposure to triethyltin in the rat. *Neurotoxicology*, 2, 97-112.


Table 1. Average Value (SE) (n=10) of Neuromuscular Measures by Dose and Time of Rats Exposed to A Single Dose of triethyltin bromide (TET)

<table>
<thead>
<tr>
<th>Dose</th>
<th>Hindlimb Grip Strength</th>
<th>Forelimb Grip Strength</th>
<th>Landing Foot Splay</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time (H)</td>
<td>Time (H)</td>
<td>Time (H)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2*</td>
<td>24*</td>
<td>168*</td>
</tr>
<tr>
<td>0.914</td>
<td>1.003</td>
<td>0.850</td>
<td>0.895</td>
</tr>
<tr>
<td>(0.050)</td>
<td>(0.075)</td>
<td>(0.055)</td>
<td>(0.067)</td>
</tr>
<tr>
<td>0.75</td>
<td>0.966</td>
<td>0.803</td>
<td>0.802</td>
</tr>
<tr>
<td>(0.042)</td>
<td>(0.065)</td>
<td>(0.041)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>1.5</td>
<td>0.865</td>
<td>0.629</td>
<td>0.619</td>
</tr>
<tr>
<td>(0.042)</td>
<td>(0.054)</td>
<td>(0.053)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>3.0</td>
<td>0.768</td>
<td>0.493</td>
<td>0.529</td>
</tr>
<tr>
<td>(0.056)</td>
<td>(0.034)</td>
<td>(0.029)</td>
<td>(0.070)</td>
</tr>
<tr>
<td>6.0</td>
<td>0.858</td>
<td>0.458</td>
<td>0.436</td>
</tr>
<tr>
<td>(0.057)</td>
<td>(0.035)</td>
<td>(0.026)</td>
<td>(0.062)</td>
</tr>
<tr>
<td>9.0</td>
<td>0.858</td>
<td>0.458</td>
<td>0.436</td>
</tr>
<tr>
<td>(0.057)</td>
<td>(0.035)</td>
<td>(0.026)</td>
<td>(0.062)</td>
</tr>
</tbody>
</table>

* Repeated ANOVA indicates a significant overall dose-effects at this time point compared with baseline

† One rat died before 168H and another had missing data (n=8).
†† One rat died before 168H (n=9).
Table 2. Toxico-diffusion model fit to hind-limb grip strength of rats exposed to triethyltin

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std Error</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_0 )</td>
<td>0.8830</td>
<td>0.0226</td>
<td>146</td>
<td>39.08</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( B )</td>
<td>-0.2138</td>
<td>0.1011</td>
<td>146</td>
<td>-2.12</td>
<td>0.0361</td>
</tr>
<tr>
<td>( C )</td>
<td>0.5934</td>
<td>0.3108</td>
<td>146</td>
<td>1.91</td>
<td>0.0582</td>
</tr>
<tr>
<td>( K_e )</td>
<td>0.0343</td>
<td>0.0031</td>
<td>146</td>
<td>11.17</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Std Error</td>
<td>Intercept (( a_i ))</td>
<td>Residual</td>
<td>0.0890</td>
<td>0.1563</td>
<td></td>
</tr>
</tbody>
</table>

Model: \( A_0 + a_i + B * d * t * \exp(-K_e * t) / (1 + C * d * t * \exp(-K_e * t)) \)

Table 3. Complementary–exponential model fit to forelimb grip strength of rats exposed to triethyltin

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std Error</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_0 )</td>
<td>1.0210</td>
<td>0.0169</td>
<td>145</td>
<td>60.11</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>( B )</td>
<td>-0.0073</td>
<td>0.0021</td>
<td>145</td>
<td>-3.44</td>
<td>0.0008</td>
</tr>
<tr>
<td>( C )</td>
<td>0.0046</td>
<td>0.0024</td>
<td>145</td>
<td>1.89</td>
<td>0.0603</td>
</tr>
<tr>
<td>( K_e )</td>
<td>0.0191</td>
<td>0.0023</td>
<td>145</td>
<td>8.26</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Std Error</td>
<td>Intercept (( a_i ))</td>
<td>Residual</td>
<td>0.0397</td>
<td>0.1759</td>
<td></td>
</tr>
</tbody>
</table>

Proportionality of Dose-specific Std Err to Control

<table>
<thead>
<tr>
<th>Dose (mg/kg)</th>
<th>Proportionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75</td>
<td>0.8817</td>
</tr>
<tr>
<td>1.5</td>
<td>0.9549</td>
</tr>
<tr>
<td>3</td>
<td>0.8031</td>
</tr>
<tr>
<td>6</td>
<td>1.4276</td>
</tr>
</tbody>
</table>

Model: \( A_0 + a_i + B * d * t * \exp(-K_e * t) / (1 + C * d * t * \exp(-K_e * t)) \)
Table 4. Quadratic polynomial model fit to landing foot splay of rats exposed to triethyltin.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std.Error</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_0$</td>
<td>76.3163</td>
<td>2.2813</td>
<td>97</td>
<td>33.45</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>$A_1$</td>
<td>-0.4423</td>
<td>0.1235</td>
<td>97</td>
<td>-3.58</td>
<td>0.0005</td>
</tr>
<tr>
<td>$B_1$</td>
<td>2.5051</td>
<td>0.3816</td>
<td>97</td>
<td>6.56</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>$C_1$</td>
<td>-0.1044</td>
<td>0.0154</td>
<td>97</td>
<td>-6.74</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Std Error</th>
<th>Intercept (a)</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12.3820</td>
<td>11.4746</td>
</tr>
</tbody>
</table>

Model: $A_0+a_i+A_1t+B_1d^*t+C_1d^*t^2$
Figure 1. Observed (top row) and predicted changes (bottom row) of hindlimb grip strength of rats exposed to a single dose of TET at 0, 0.75, 1.5, 3.0, and 6.0 mg/kg, respectively. Dose groups are arranged in separate panels in an increasing order from left to right. Predicted changes are based on a toxico-diffusion model fitted to the observed changes. Time scale is in logarithm.
**Figure 2.** Observed (top row) and predicted changes (bottom row) of forelimb grip strength of rats exposed to a single dose of TET at 0, 0.75, 1.5, 3.0, or 6.0 mg/kg, respectively. Dose groups are arranged in separate panels in an increasing order from left to right. Predicted changes are based on a complementary-exponential model fit to the observed data. Time scale is in Logarithm.
Figure 3a. Observed changes of landing foot splay of rats exposed to a single dose of TET at 0, 0.75, 1.5, 3.0, or 6.0 mg/kg, respectively. Dose groups are arranged in separate panels in an increasing order from left to right. Observed values (symbols) of the same rat are linked through lines. Time scale is in Logarithm.

Figure 3b. Model-fitted changes of landing foot splay of rats exposed to a single dose of TET at 0, 0.75, 1.5, 3.0, or 6.0 mg/kg, respectively. Dose groups are arranged in separate panels in an increasing order from left to right. Data at 168 hours were excluded to achieve model fitting by the polynomial function. Time scale is in Logarithm.
**Figure 4.** Individual rat’s changes in hindlimb grip strength after a single dose of TET. Each row represents a dose group, from top to bottom, at 0, 0.75, 1.5, 3.0, and 6.0 mg/kg, respectively. Dotted line (labeled “ID”) is individual rat’s trajectory under the toxico-diffusion model; solid line (labeled “fixed”) is the dose-group average trajectory; circles are the individual rat’s observed values.
Figure 5. Standardized residuals versus fitted values under the toxico-diffusion model fit to hindlimb grip strength of rats exposed to TET. Residuals are, from left to right, grouped in panel in increasing dose level.
Figure 6. Individual rat’s changes in forelimb grip strength after a single dose of TET. Each row represents a dose group, from top to bottom, at 0, 0.75, 1.5, 3.0, and 6.0 mg/kg, respectively. Dotted line (labeled “ID”) is individual rat’s trajectory under the complementary-exponential model; solid line (labeled “fixed”) is the dose-group average trajectory; circles are the individual rat’s observed values.
Figure 7. Standardized residuals versus fitted values under the complementary-exponential model fit to forelimb grip strength of rats exposed to TET. Residuals are, from left to right, grouped in panel in increasing dose level.
Figure 8. Individual rat’s changes in landing foot splay after a single dose of TET. Each row represents a dose group, from top to bottom, at 0, 0.75, 1.5, 3.0, and 6.0 mg/kg, respectively. Dotted line (labeled “ID”) is individual rat’s trajectory under a polynomial model; solid line (labeled “fixed”) is the dose-group average trajectory; circles are the individual rat’s observed values.
Figure 9 Standardized residuals versus fitted values under the polynomial model fit to landing foot splay of rats exposed to TET. Residuals are, from left to right, grouped in panel in increasing dose level.