©2010

Susan Huyck

MODELING TIMES OF MAXIMUM BIOMARKER EXCRETION

By SUSAN HUYCK

A dissertation submitted to the

School of Public Health

University of Medicine and Dentistry of New Jersey

and the

Edward J. Bloustein School of Planning and Public Policy

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

for the degree of

Doctor of Public Health
UMDNJ-School of Public Health

Awarded jointly by these institutions and

Written under the direction of

Professor Pamela Ohman-Strickland

And Approved by

_____

_____

_____

_____

_____

Piscataway/New Brunswick, New Jersey

January, 2010

ABSTRACT OF THE DISSERTATION

Modeling Times of Maximum Biomarker Excretion

by SUSAN HUYCK

Dissertation Director:

Pamela Ohman-Strickland

Analysis of biomarkers to detect levels of chemical exposure in humans is an important risk evaluation tool. For example, urine biomarkers such as 1-aminopyrene can be used to assess exposure levels to diesel exhaust (DE), an important public health concern. Toxic chemicals contained in DE and DE particles have demonstrated genotoxic and carcinogenic properties in experimental animals. A recent experiment evaluating the urine concentration of DE biomarkers was the impetus for this dissertation. One goal of the experiment, and the focus of this dissertation, was to characterize the excretion time course of the biomarker 1-aminopyrene. The times of maximum concentration in plasma or maximum excretion in urine have been typically summarized using non-parametric or asymptotic techniques based on individual subject-level values; however, there is limited information addressing confidence interval generation when sparse subject-level samples requiring population-modeling approaches are present. Therefore, there was a need to generate and evaluate an appropriate confidence interval approach when sparse sampling is present.

Pharmacokinetic (PK) modeling was used to fit a standard one-compartment urine excretion model to the data for estimation of the time of maximum excretion. Several

variations of the PK model were explored and a model based on cumulative excretion rates was selected. Several statistical techniques for modeling PK data and calculating confidence intervals for the time of maximum excretion were compared including confidence intervals based on the first and second order delta methods, derived for this dissertation.

A comparison of confidence interval methods showed that when using: (1) within-subject Tmax values, coverages obtained using the non-parametric method were highest and often provide coverages close to the nominal 95% level; and (2) population-average Tmax values, confidence intervals generated using the first-order delta method provided the highest coverages, at approximately 93% when numerical approximation estimation methods were used. Subject response profiles for the 1-aminopyrene biomarker data were varied and led to a hypothesis that a mixture of more than one distribution of profiles may be present. Future exploration with data collected for more than 24-hours would be needed to further explore this hypothesis fully.

ACKNOWLEDGMENT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Chapter 1

Introduction and Overview

This dissertation was motivated by a recent experiment designed to evaluate the urine concentration of diesel-exhaust biomarkers. In this two-period, cross-over design study, subjects were exposed to controlled levels of diesel-exhaust in one period and clean air in the other period. Spot urine samples were collected immediately prior to exposure and at the subject's convenience for 24-hours following exposure and tested for diesel-exhaust biomarker levels. This dissertation addresses one research goal of the experiment; which was to characterize the excretion time course of the biomarker, 1-aminopyrene through point estimation and confidence intervals in order to optimize sampling times in future studies.

This study was undertaken because exposure to diesel exhaust constitutes a public health concern as DE contains a large suite of toxic chemicals including nitro-polycyclic aromatic hydrocarbons (nitro-PAHs) [Bond et al., 1986, Howard et al., 1995]. Many nitro-PAHs (e.g., nitropyrenes) have been shown to have genotoxic and carcinogenic properties in experimental animals [Beland and Kadlubar, 1990, IARC Monographs, Tokiwa and Ohnishi, 1986]. Concerns about the public health risks resulting from inhalation exposures to nitropyrenes have inspired numerous studies to further the scientific knowledge on the deposition, metabolism and excretion of nitropyrenes [Bond et al.,1986, Dutcher et al., 1985, Sun et al., 1983]. A recent study found that concentrations of nitro-PAHs were 3 orders of magnitude higher in DE particles than in urban ambient particles, further supporting the notion that nitro-PAHs are highly specific

to diesel combustion emissions [Bamford and Baker, 2003]. The most abundant nitro-PAH in DE particles is 1-nitropyrene (1-NP). 1-NP is metabolized to 1-aminopyrene (1-AP) and has been measured in the urine of DE-exposed miners [Seidel et al., 2002]. However, little is known about detailed timing of 1-AP excretion following a spiked DE exposure.

The standard approach used to evaluate the pharmacokinetic properties of absorption, distribution, metabolism and excretion related to a chemical entity involves the collection and analysis of serial measurements of biological specimens such as plasma or urine from a sample of individuals within a specified time period following exposure to the chemical entity. In general, estimates of absorption and excretion rates along with maximum concentration and time to maximum concentration are typical parameters of interest to the researcher. The objective is to characterize the time and concentration relationship as determined by pharmacokinetic processes within the body for individual subjects as well as the population of subjects as a whole. A typical feature of the data is that the non-linear relationship between the response and time will often vary between individuals. Thus, non-linear mixed effects models are applied in order to handle repeated measures data and allow for flexible variance-covariance structures (Davidian and Giltinan 2003). These models assume that the form of the intra-subject model which relates the response to time is common to all subjects, but aim to understand the "typical" population pharmacokinetic behavior for the outcome of interest, (e.g. maximum concentration) as well as the degree to which this behavior varies across individuals.

In order to characterize the time trend of the urine DE biomarker, 1-AP, this dissertation will address the unique aspects of spot sampled urine pharmacokinetic modeling, evaluate several data modeling approaches and compare established confidence interval methods along with a "new" method based on the delta method. The EM algorithm is applied to the 1-AP data to evaluate a mixture distribution.

This dissertation contains 6 chapters. Chapter 2 addresses the selection of an appropriate pharmacokinetic model for the urine excretion data. Using time and excretion data from urine samples as opposed to time and concentration data from plasma samples can present certain challenges to the pharmacokinetic modeling process. While plasma samples are more commonly collected in pharmaceutical studies, urine sample collection is much less invasive and do not require subject presence at the study site. However, urine data represents concentration of a chemical substance accumulated since the previous void and not the concentration at a particular time point, as with plasma data. In addition, spot urine sampling results in random and sometimes sparse information for individual subjects. The level of information available for individual subjects can have a direct effect on the nonlinear modeling approach used.

Chapter 3 explores the varied response profiles observed in the DE experiment and hypothesizes potential reasons for these differences among the study volunteers using the experimental DE data as well as several simulated data sets. Chapter 4 compares parametric and non-parametric confidence interval techniques as well as non-compartmental and compartmental methods using nonlinear modeling. Depending on the amount of information available for individual subjects, various modeling approaches may be used to estimate the time of maximum excretion, either at the individual subject-

level and/or for the population as a whole. When insufficient individual subject-level information is available, population-average pharmacokinetic modeling techniques are used to estimate pharmacokinetic parameters of interest. However, standard errors for derived parameters such as the time of maximum excretion are not readily available for calculation of confidence intervals. A "new" approach based on the delta method is introduced in this chapter. Data representing homogeneous results from a theophylline plasma data set, heterogeneous results from the DE 1-AP experiment and simulated data are used to demonstrate the confidence interval techniques for the various modeling approaches.

Chapter 5 explores a hypothesis generated in Chapter 3. One potential reason for the wide variety of observed response profiles is that a mixture of more than one distribution may be present in the data. Biological explanations to support this hypothesis include potential differences among subjects in metabolic capacities (polymorphisms) [Saito, et al. 1984] or differences in exposures based on respiration rates, etc. The Expectation-Maximization (EM) algorithm [Dempster, Laird, Rubin 1977] is used with individual subject-level pharmacokinetic parameter estimates to explore a mixture of two multivariate normal distributions. An overall discussion of the results and suggestions for future work can be found in Chapter 6.

References

Bamford, H.; Baker, J. Nitro-polycyclic aromatic hydrocarbon concentrations and sources in urban and suburban atmospheres of the Mid-Atlantic region. *Atmos. Environ. 2003. 37(15),* 2077-2091.

Beland, F. A.; Kadlubar, F. F. Metabolic Activation and DNA Adducts of Aromatic Amines and Nitroaromatic Hydrocarbons. In *Carcinogenesis and Mutagenesis I* ; Cooper, C.S.; Grover, P.L., Eds.; Springer-Verlag: New York, 1990, 267-325.

Bond, J. A., Sun, J. D., Medinsky, M. A., Jones, R. K. , Yeh, H. C. (1986). Deposition, Metabolism, and Excretion of 1-[^{14}C]Nitropyrene and 1-[^{14}C]Nitropyrene Coated on Diesel Exhaust Particles as Influenced by Exposure Concentration. *Toxicology and Applied Pharmacology.* 85, 102-117.

Dempster, A., Laird, N., Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B,* 39 (1) 1-38.

Dutcher, J. S., Sun, J. D., Bechtold, W. E., Unkefer, C. J. (1985). Excretion and Metabolism of 1-Nitropyrene in Rats after Oral or Intraperitoneal Administration. *Toxicological Sciences*, 5 (2), 287-296.

Howard, P. C., Consolo, M. C., Dooley, K. L., Beland, F. A. (1995). Metabolism of 1-Nitropyrene in Mice: Transport Across the Placenta and Mammary Tissues. *Chemico-Biological Interactions,* 95, 309-325.

IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. http://monographs.iarc.fr/ENG/Monographs/vol46/volume46.pdf (accessed July 30, 2008). Volume 46 Diesel and Gasoline Engine Exhausts and Some Nitroarenes.

Saito, K.; Kamataki, T.; Kato, R. Participation of Cytochrome P-450 in Reductive Metabolism of 1-Nitropyrene by Rat Liver Microsomes. *Cancer Res*. 1984, 44, 3169-3173.

Seidel, A.; Dahmann, D.; Krekeler, H.; Jacob, J. Biomonitoring of polycyclic aromatic compounds in the urine of mining workers occupationally exposed to diesel exhaust. *Int. J. Hyg. Environ. Health.* 2002. *204,* 333-338.

Sun, J. D., Wolff, R. K. , Aberman, H. M. , McClellan, R. O. (1983). Inhalation of 1-nitropyrene Associated with Ultrafine Insoluble Particles or as a Pure Aerosol: A Comparison fo Deposition and Biological Fate. *Toxicology Applied Pharmacology,* 69 (2) 185-198.

Tokiwa, H.; Ohnishi, Y. Mutagenicity and carcinogenicity of nitroarenes and their sources in the environment. *Crit Rev Toxicol.* 1986. *17(1),* 23-60.

Chapter 2

Modeling Pharmacokinetic Urine Concentration Profiles of Diesel Exhaust Biomarker 1-aminopyrene

Abstract

Residual analysis techniques were used to compare three pharmacokinetic models applied to urine excretion data for a Diesel Exhaust biomarker, 1-aminopyrene (1-AP). Variations in subject response profiles prompted an evaluation of several models to find the optimal fit using residual analysis. Graphical and statistical methods were used to evaluate the underlying assumptions of the residual distribution including tests for normality, location, and homogeneity of variance. The urine excretion pharmacokinetic model, based on first-order principles, was examined via a rate of excretion model, cumulative excretion model, and change in cumulative excretion model. The cumulative excretion model resulted in a better fit, providing residuals that more closely adhered to normality, location, and variance homogeneity assumptions and providing unbiased estimates, on average, across concentration levels.

Introduction

Pharmacokinetic models are used to describe the time and concentration profiles of drugs and other chemicals following exposure. Identifying the most appropriate model is an important step in the process of pharmacokinetic data analysis. While the typical pharmacokinetic analysis is based on data from plasma samples, urine excretion samples represent an important modeling arena. Urine collection is less invasive than plasma sampling and does not require subject presence at the clinic site at the time of sample

collection as voids can be collected and retained under appropriate storage conditions for later delivery to the clinic site. Collection of spot (untimed) urine samples is an important tool for biomonitoring public health parameters such as exposure to diesel exhaust (DE) and DE particles. Simplification of the timing of samples would bring efficiency to the biomonitoring process. Increasing the efficiency not only would reduce costs but also encourage subject compliance by decreasing the number of samples required. This chapter uses a residual analysis approach to compare three pharmacokinetic models used to describe urine excretion data from a DE experiment and identifies the model that best fit the data for use in future work.

Exposure to DE represents a public health concern due to the environmental contaminants contained in DE and DE particles. These environmental contaminants are nitropolycyclic aromatic hydrocarbons (nitro-PAHs) which are found in a variety of sources including diesel exhaust (DE) and coal combustion fly ash (Howard, et al. 1995 and Bond, et al. 1986). A sub-class of nitro-PAHs, the nitropyrenes, has been shown to have genotoxic and carcinogenic properties in experimental animals (IARC Monographs 1989, Tokiwa & Ohnishi 1986, Beland & Kadlubar, 1990). Concerns about the public health risks resulting from inhalation exposures to nitropyrenes have inspired numerous studies to further the scientific knowledge on the deposition, metabolism and excretion of nitropyrenes (Bond, et al 1986, Sun, et al. 1983, Dutcher, et al. 1985). Results of these studies have shown that nitro-PAHs are rapidly absorbed and metabolized. The main route of excretion following inhalation is through the feces and the urine as determined by ^{14}C-radiolabeled exposure, with about 2 times more being excreted through the feces

than the urine. Half-times for elimination of ^{14}C in urine and feces range from 15 to 20 hours (Bond, et al 1986).

A recent study found that concentrations of nitro-PAHs were 3 orders of magnitude higher in DE particles than in urban ambient particles (Bamford 2003). The most abundant nitro-PAH in DE particles was 1-nitropyrene (1-NP). 1-NP is metabolized to 1-aminopyrene (1-AP) which has been measured in the urine of DE-exposed miners (Seidel 2002). Knowledge of the time-course of excretion following DE exposure, presently unknown for 1-AP, would allow for optimization of the sampling times and more efficient biomonitoring.

Using time and concentration data from urine samples as opposed to plasma samples can present certain challenges to the pharmacokinetic modeling process. Serial plasma samples are a commonly used source for measuring concentrations over time. Each plasma sample represents the concentration of the substance being measured at that point in time. In contrast, a urine sample collected at a particular time point represents the cumulative amount excreted since the previous void. Note that while plasma sample collection times are typically standardized across individual subjects within a study; spot urine sampling results in random and sometimes sparse collections. The collection of samples at random times can result in a variety of levels of information being presented by an individual subject. For example, a subject with more frequent voids will provide more specific information regarding the time of the peak excretion compared to a subject with few voids.

Pharmacokinetic modeling techniques are well-suited to the characterization of time and concentration relationships of chemical entities which have been absorbed,

metabolized and excreted by the body. Time to peak concentration for a chemical entity can be approximated by modeling the rate of urinary excretion or cumulative amount excreted. Under the following pharmacokinetic assumptions: 1) presence of a one-compartment model; 2) first-order input; 3) first-order elimination; and 4) single route of absorption (in this case, inhalation); a standard pharmacokinetic model for the total amount of a substance excreted in urine is given by (Shargel & Yu, 1993):

$$D_u(t) = \frac{FKeKaD}{Ka-K}\left(\frac{1}{K} - \frac{1}{Ka}\right) + \frac{FKeKaD}{Ka-K}\left[\frac{1}{Ka}e^{-Kat} - \frac{1}{K}e^{-Kt}\right]$$

A modified version of this model was used for the 1-AP data to reflect the fact that 1-AP is a metabolite of 1-NP and not directly absorbed through inhalation. Assuming a first-order rate of metabolism, we substitute the *Ka* term with *Km* to reflect this modification:

$$D_u(t) = \frac{FKeKmD}{Km-K}\left(\frac{1}{K} - \frac{1}{Km}\right) + \frac{FKeKmD}{Km-K}\left[\frac{1}{Km}e^{-Kmt} - \frac{1}{K}e^{-Kt}\right] \qquad (2.1)$$

where $\dfrac{FKeKmD}{Km-K}\left(\dfrac{1}{K} - \dfrac{1}{Km}\right)$ and $\dfrac{FKeKmD}{Km-K}$ are constants with *F* representing the fraction of drug absorbed, *Ke* equal to the first order renal excretion constant, *Km* equal to the first order rate constant for metabolism, *K* equal to the first order rate constant for elimination, and *D* is the amount of drug, or chemical entity exposed to the subject. Figure 2.1 below illustrates the typical shape describing cumulative urinary biomarker excretion over time, following a single exposure.
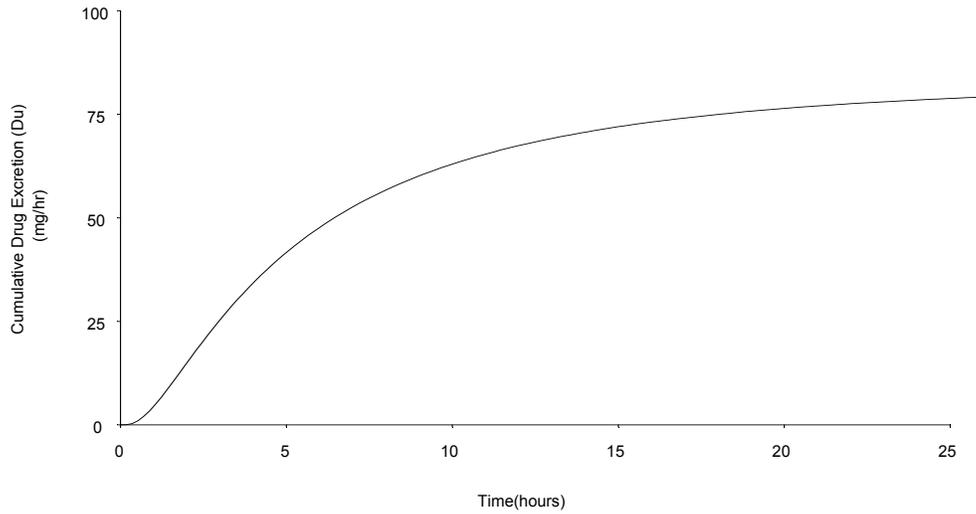
Figure 2.1: Illustration of Cumulative Excretion

Taking the first derivative of equation (2.1) with respect to time, yields the rate of urinary excretion of a substance over time.

$$\frac{\partial Du(t)}{\partial t} = \frac{\partial}{\partial t}\left\{\frac{FKeKmD}{Km-K}\left(\frac{1}{K}-\frac{1}{Km}\right)+\frac{FKeKmD}{Km-K}\left[\frac{1}{Km}e^{-Kmt}-\frac{1}{K}e^{-Kt}\right]\right\}$$

$$=\left(\frac{FKeKmD}{Km-K}\right)\times\frac{\partial}{\partial t}\left[\frac{1}{Km}e^{-Kmt}-\frac{1}{K}e^{-Kt}\right]$$

$$=\frac{FKeKmD}{Km-K}\left(e^{-Kt}-e^{-Kmt}\right) \tag{2.2}$$

Figure 2.2, below, demonstrates a typical first-order absorption and elimination model for urinary excretion rate.
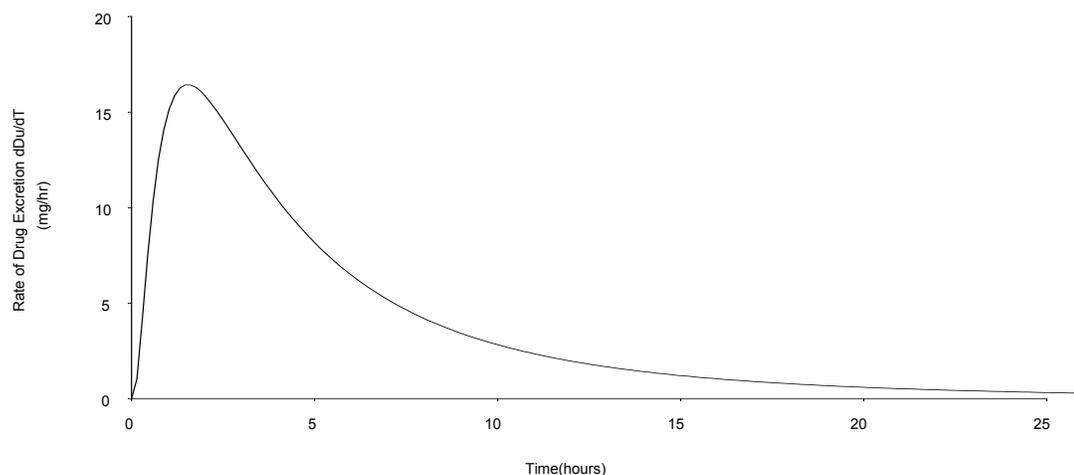
Figure 2.2: Illustration of Rate of Excretion

In a recent experiment, spot urine samples representing partial voids were collected from healthy volunteers during a 24-hour period following a brief controlled exposure to DE. The overall goal of the experiment was to characterize the time course of 1-AP excretion in order to optimize sampling times in future studies.

Visual review of subject concentration profiles revealed that a large variety of data patterns were present.   Three modeling approaches were considered for fitting the concentration profiles:

(1) a model based on the rate of urine excretion over time;

(2) a model based on the cumulative concentration over time; and,

(3) a model based on the change in cumulative concentration over time.

This chapter uses a residual analysis approach to compare three pharmacokinetic models used to describe urine excretion data from a Diesel Exhaust experiment and identifies the model that best fits the data for use in future analyses.

Materials and Methods

Diesel Exhaust Experiment

In a recent study of DE exposure, biomarkers were measured in 55 normal volunteers using spot urine samples collected over a 24-hour period. In a two-period crossover design, subjects underwent one hour of controlled exposure to either filtered-air (clean air, "CA") or diluted DE (300 µg/m3 as PM 2.5) in a randomized order on separate mornings at least one week apart. Randomized subjects reported to the Clinical Center prior to exposure and provided one urine sample. Inhalation exposure to either CA or DE occurred at the Controlled Environment Facility (CEF) at the Environmental and Occupational Health Sciences Institute (EOHSI). Following the one-hour exposure subjects were instructed to collect samples of all urine voids that occurred prior to their return to the Clinical Center 24 hours later. Urine samples were collected in 50-ml sterile collection cups. Subjects were provided a cooler for immediate storage of urine samples prior to their return to the Clinical Center. All recruitment and testing procedures were reviewed and approved by the Institutional Review Board of UMDNJ.

Measurements of urine volume voided were not available due to the partial void sampling method. Concentration is a function of rate of excretion, volume voided and time. In order to accommodate partial-void sampling, generally used for biomonitoring purposes, methods for standardization of the biomarker concentration have been developed. The routine method used for this standardization is an adjustment based on urinary creatinine concentration where the biomarker is reported as a weight per gram of creatinine. This works well when making comparisons within individuals because the intra-individual variation in amount of creatinine excreted per day is relatively low. (Barr

2005). Creatinine-standardized concentrations of urine biomarkers have been calculated for the data in this study so that comparisons within and between subjects could be made that would reflect the amount of 1-AP excreted accounting for differences in volume of output.

Details on the CEF, personal air sampling, and the assay methods for measuring 1-AP and creatinine have been provided in a prior publication by Laumbach, et al. 2009.

Statistical Methods

Approximately 64 of the 372 (17%) total DE and 51 of 356 (14%) total CA urine samples were considered invalid based on creatinine levels that were either missing, too low (<30 mg/dL) or too high (>300 mg/dL). This led to the development of rules for inclusion of a subject's data for model evaluation. These rules were similar to those used in a previous publication (Laumbach, et al. 2009). Specifically, to be included in the analysis data set, a subject must have provided at least two valid urine samples post exposure and have had no more than one invalid sample in a row in each of the CA and DE arms. As a sensitivity analysis, the data set which included all subjects and all samples was also examined.

Only the DE exposure data was used in the statistical analysis to determine the best fitting model for characterization of the time-concentration profile following DE exposure, because this is the primary focus of this research. Creatinine-standardized concentrations of 1-AP were modeled using SAS Proc NLMixed for nonlinear mixed effects modeling. Specifically, if y^*_{ij} is standardized value for the ith subject ($i=1,\ldots, m$) at the jth time point, ($j=1,\ldots, n_i$) then $y_{ij} = \sum_{j=2}^{n_i} (t_{ij} - t_{i,j-1}) \times y^*_{ij}$ is the cumulative

concentration for subject i at time j where j are the post-exposure time points and $y_{ij}=y^*_{i1}$ at $j=1$. The following three modeling approaches were compared:

Model 1: the rate of urine excretion over time denoted by

$$E\left(y^*_{ij}\right) = \frac{F_i Ke_i Km_i D_i}{Km_i - K_i}\left(e^{-K_i t_{ij}} - e^{-Km_i t_{ij}}\right);$$

Model 2: the cumulative concentration over time denoted by

$$E\left(y_{ij}\right) = \frac{F_i Ke_i Km_i D_i}{Km_i - K_i}\left(\frac{1}{K_i} - \frac{1}{Km_i}\right) + \frac{F_i Ke_i Km_i D_i}{Km_i - K_i}\left[\frac{1}{Km_i}e^{-Km_i t_{ij}} - \frac{1}{K_i}e^{-K_i t_{ij}}\right]$$

Model 3: the change in cumulative concentration over time denoted by

$$E\left(y^c_{ij}\right) = \left(y_{ij} - y_{i,j-1}\right)/\left(t_{ij} - t_{i,j-1}\right).$$

Time was defined as the elapsed time in hours since the end of the one-hour controlled DE exposure. The parameters planned to be estimated in the nonlinear mixed model included F, Ke, Km, and K – the fraction of DE absorbed, the renal excretion rate, the 1-AP metabolism rate, and the elimination rate. D, the amount of DE exposed to the subject, was not a parameter to be estimated because all subjects were exposed to the same level of controlled DE.

The statistical plan specified that both the First Order Taylor Series expansion and the Laplace methods would be used to approximate parameter estimates. These two approximation methods along with the three modeling approaches listed above were to be compared by examining the residuals to determine which model provided the best fit to the data. In nonlinear modeling as with linear regression analysis, residuals were expected to follow a normal distribution with a mean of zero and a variance that was homogenous across the predictor variable(s), e.g. in this application, sample time. A well-fitting model would have residuals that not only possessed these characteristics, but

also would have been smaller in magnitude indicating that the predicted values were closer to the observed data.   Residuals were calculated as the difference between the predicted values and the observed data. Residuals were compared graphically and with statistical tests for normality, location (to ascertain whether the values were centered around zero) and homogeneity of variance across time.   Normality was assessed using the Shapiro-Wilk and Kolmogorov-Smirnov test, location with the Sign-rank test, and homogeneity of variance with Levene's test and the Brown-Forsythe test, where Levene's test is based on the mean and the Brown-Forsythe (BF) test is based on the median and provides a degree of robustness against data that may not follow a normal distribution.

This study was a 2-period crossover design.   The DE exposure results were generated in both periods, with approximately half the subjects being exposed to DE in the first period and the other half being exposed in the second period. Prior to pooling the DE data across all subjects for the nonlinear mixed effects modeling comparisons, an ANOVA model for crossover experiments was used to determine whether there were any statistically significant effects due to period or sequence. The response variable used to evaluate sequence and period effects was the time-weighted average concentration calculated as $TW_{hi} = \sum_{j=2}^{n_i} y_{hij}^* \left(t_{hij} - t_{hij-1}\right) \Big/ \sum_{j=2}^{n_i} \left(t_{hij} - t_{hij-1}\right)$ where TW_{hi} is the time-weighted average for subject i following exposure to DE or CA, and h =DE or CA.

Results

Fifty-five healthy volunteers participated in this DE exposure study. This group consisted of 33 males and 22 females with an average age of 24.8 years (Table 2.1). Fourteen subjects had incomplete or invalid data following the DE exposure arm and one

subject discontinued prior to completing the study. These 15 subjects were excluded from the primary model evaluation analysis subset. The 40 analyzed subjects consisted of 25 males and 15 females with an average age of 24.0 years.

Table 2.1: Summary of Demographic Characteristics

	All Randomized Subjects (n=55)	Evaluable Subset (n=40)
Gender (n,%)		
Male	33 (60%)	25 (63%)
Female	22 (40%)	15 (37%)
Age (years)		
Mean (Std Dev)	24.8 (6.59)	24.0 (5.54)
Range	19 - 44	19 - 43

Eighty-three per cent (37/40) of the evaluable subjects provided 5 or more valid post-DE exposure urine samples with the majority (70%) of subjects providing between 5 and 7 valid samples, details are provided in Table 2.2.

Table 2.2: Number of Valid Post-DE Exposure Urine Samples in the Evaluable Data Set

Number of Samples	Number of Subjects	Percent	Cumulative Frequency	Cumulative Percent
3	1	2.5	1	2.5
4	5	12.5	6	15.0
5	9	22.5	15	37.5
6	11	27.5	26	65.0
7	8	20.0	34	85.0
8	6	15.0	40	100.0

Scatter plots with lines indicating the time trend for each of the evaluable subjects creatinine-standardized 1-AP concentration following both DE and CA exposure are

provided in Figure 2.3. These graphs show a wide range of response patterns between individual subjects and indicate that a single population-level model may not be sufficient to describe the data. In these graphs the y-axis was allowed to vary between subjects to maximize the visualization of the curve shapes, the solid lines indicate the 1-AP concentrations following DE exposure and the dotted lines indicate the levels after the CA exposure.
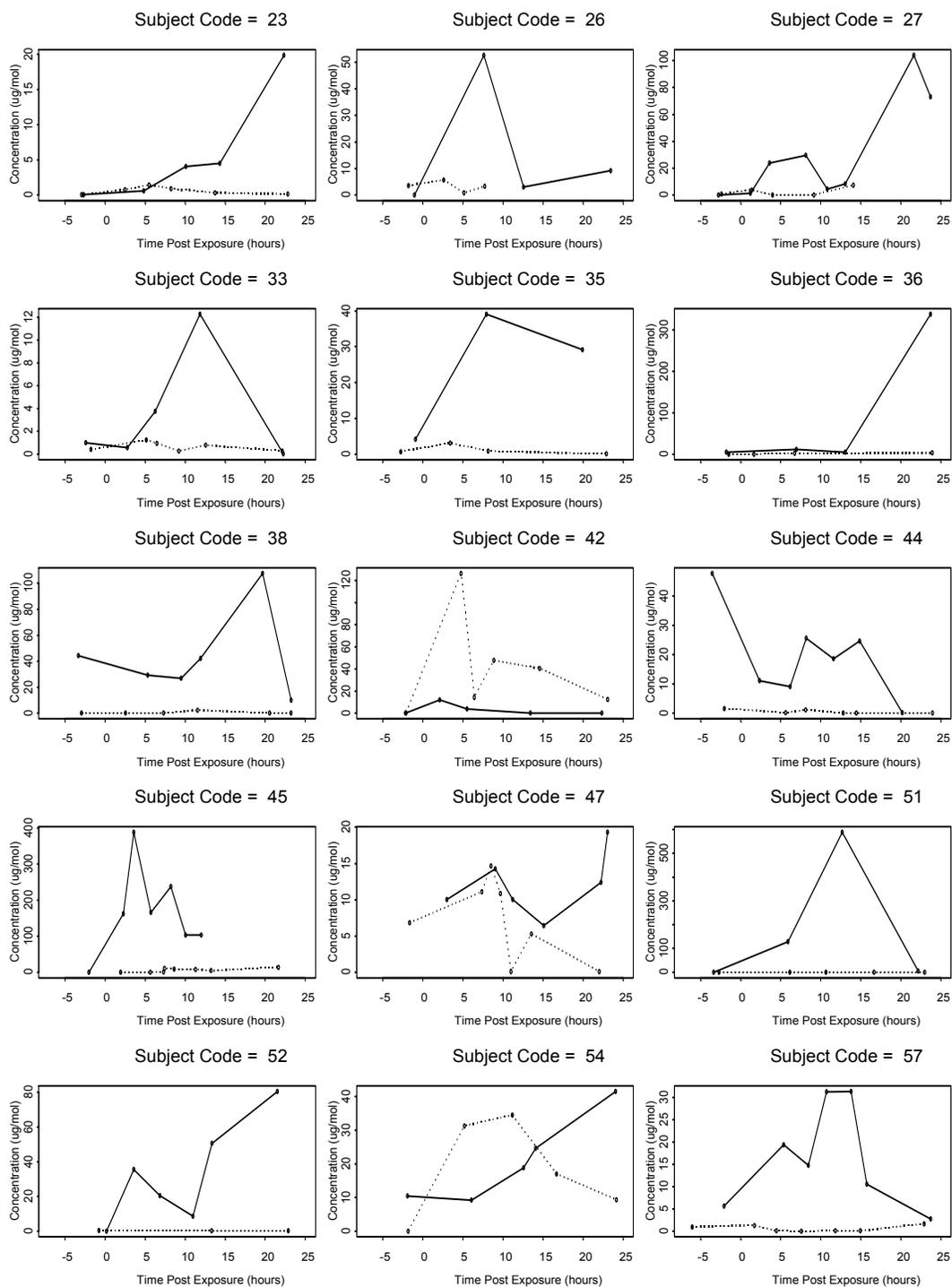
Figure 2.3: Standardized 1-AP Concentration (µg/mol) by Time and Subject
——— **Diesel exposure, - - - - - Clean air exposure**
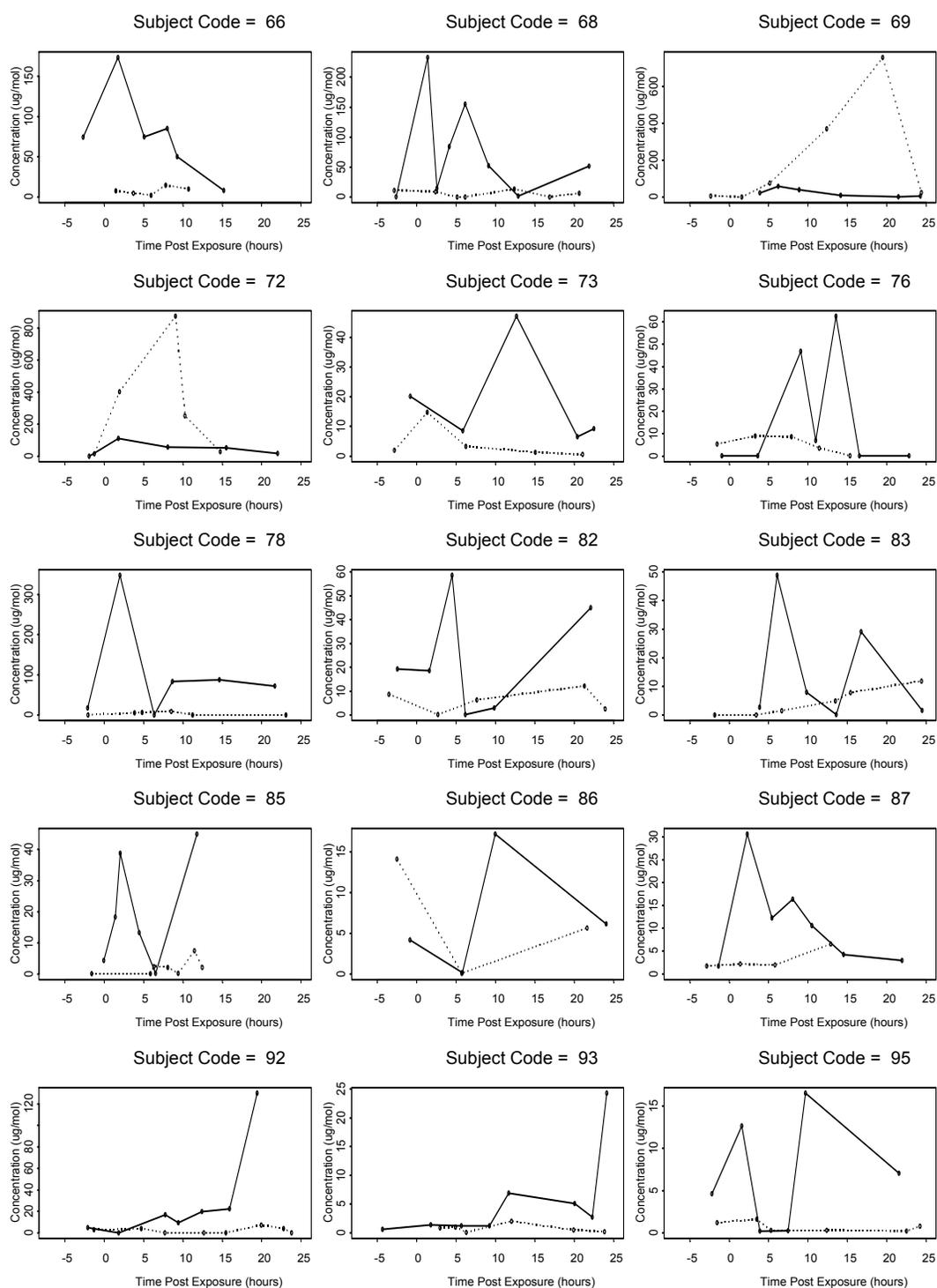
Figure 2.3: Standardized 1-AP Concentration (μg/mol) by Time and Subject
——Diesel exposure, - - - - - Clean air exposure
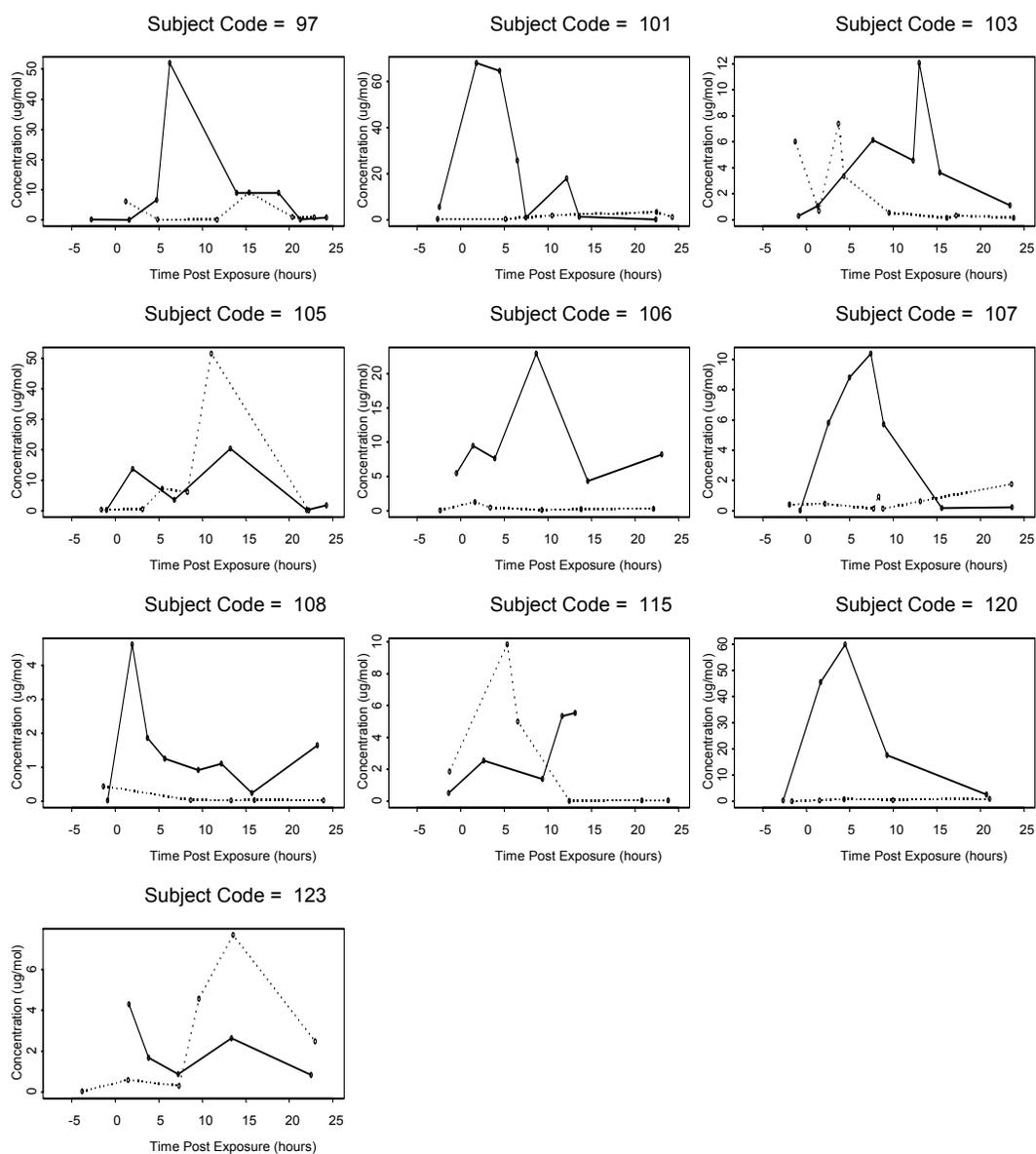
Figure 2.3: Standardized 1-AP Concentration (μg/mol) by Time and Subject
———Diesel exposure, - - - - - Clean air exposure

The standardized 1-AP concentration levels after DE exposure are shown for all 40 evaluable subjects, collectively, in the box plots in Figure 2.4 where the time point values have been rounded to the nearest hour. The shaded boxes in the plots show the range of concentrations that fall within the 25th and 75th percentiles, whiskers extend out to values that fall outside the 25th - 75th percentile but within 1.5 times the inter-quartile range, values beyond this range are denoted by the '∗' symbol. The data distribution demonstrates the presence of a number of values higher than 1.5 times the inter-quartile range indicating a wide variety of peak concentration levels.
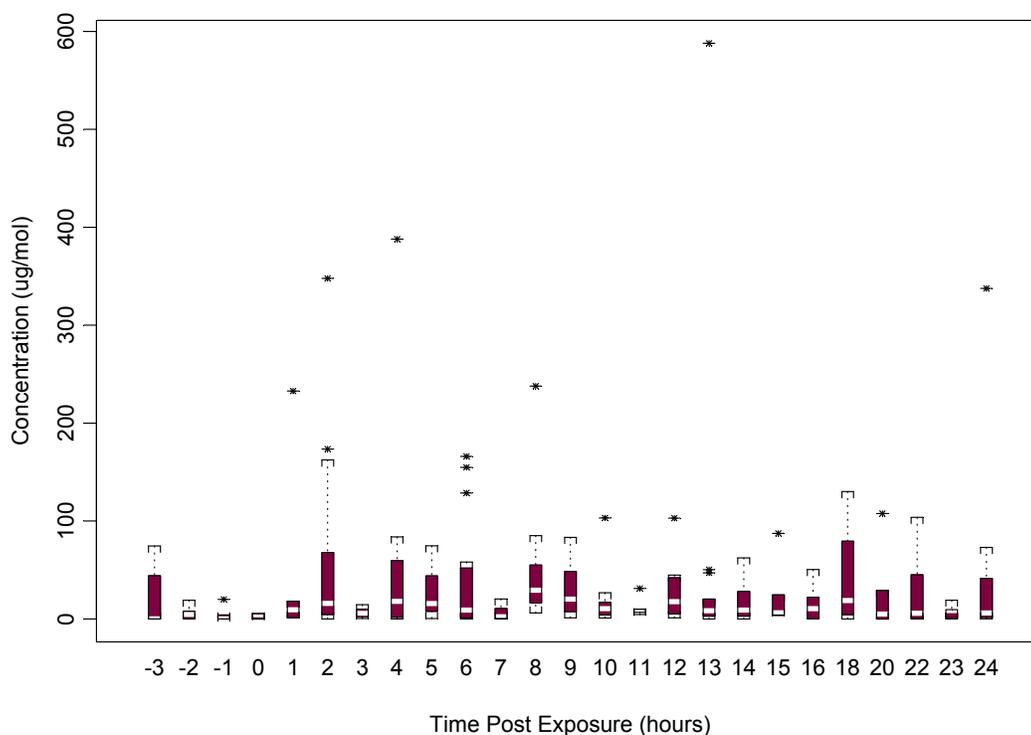
Figure 2.4: Box Plot of 1-AP Concentration by Time for DE Exposure (Time was rounded to the nearest hour. Hours with few samples were combined, e.g. hours 17-19 were combined and plotted at hour 18.)

Results of the ANOVA for the crossover design showed that there were no statistically significant period (p=0.19) or sequence effects (p=0.31), indicating that the DE exposure data could be pooled into one analysis set.

Numerous attempts at fitting population-level models using both First Order and Laplace approximation methods failed to result in model convergence for all 3 of the planned approaches; rate of excretion, cumulative excretion and change in cumulative excretion. Variations included changing the starting parameter values and identification of random terms. Population-level models converge when the concentration profiles from the individual subjects follow similar patterns. Due to the diversity of individual responses (displayed in Figures 2.3 and 2.4), no model that described all the subjects with a single set of parameters could be identified. However, it was possible to fit individual subject-level models for the three modeling approaches for most of the subjects. The Laplace method was the only method used to fit the individual subject-level data due to a software limitation with the First Order method that required identification of at least one random effect in order to run. In nonlinear mixed modeling, random effects are identified at the population level and not at the subject level, therefore there were no random effects and the First Order method could not be used.

In general the three modeling approaches were similar in their ability to achieve model convergence for the individual subjects. Only the model based on the cumulative change in concentration over time (Model 3) failed to converge for two subjects (#42 and 44) resulting in predicted concentration values of zero at all time points.

A residual analysis based on standardized residuals was used to evaluate the fit of the three models. Residuals were standardized by dividing each value by the standard

deviation of the relevant model residuals to account for the differences in magnitude of concentration between the rate data and the cumulative concentration values. Four sets of plots comparing residuals are presented in Figures 2.5 through 2.8; the first two show the distribution of the residuals with histograms and quantile-quantile (QQ) plots, the third shows the residuals versus the 1-AP concentrations, and the fourth shows the residuals over time. In each figure, Model 1 residuals are displayed in the top graph, Model 2 in the middle, and Model 3 at the bottom, where Model 1 represents the fit based on the rate of urine excretion over time, Model 2 was based on the cumulative concentration over time and Model 3 was based on the change in cumulative concentration over time. Results of the statistical tests for normality, location, and homogeneity of variance are presented in Table 2.3.
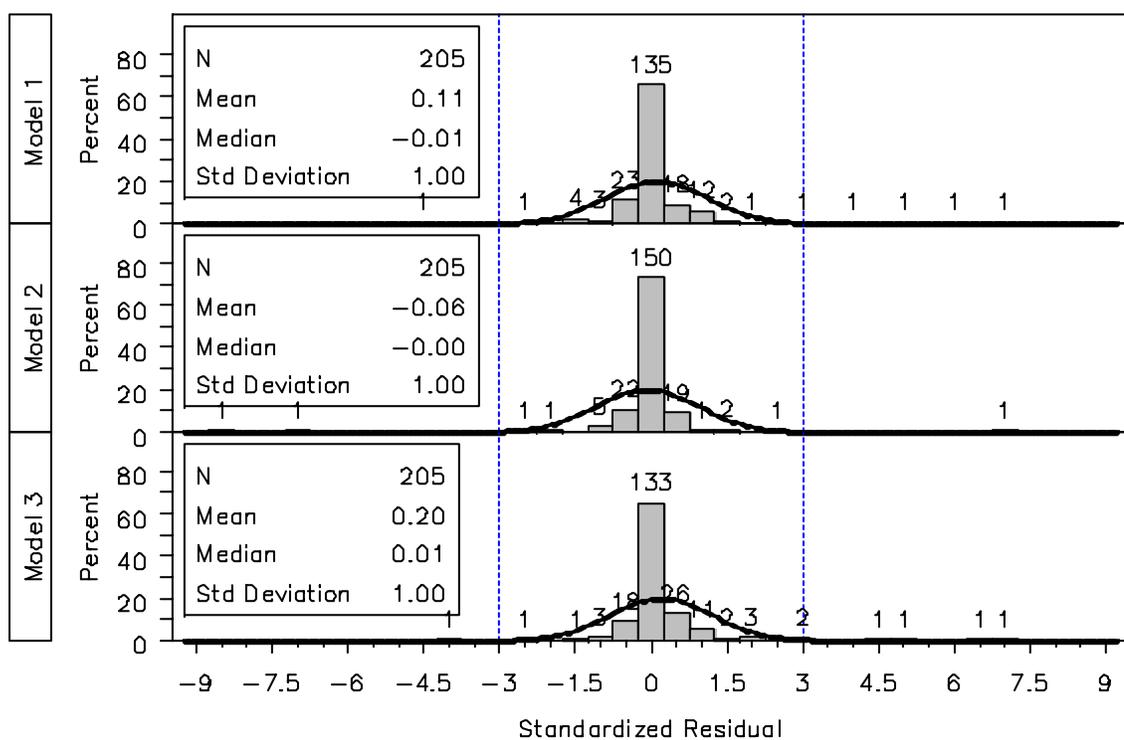
Figure 2.5: Standardized Residual Distribution

Normality

The histograms in Figure 2.5 display bars depicting the proportion of standardized residuals in the relevant value range where the numbers at the top of the bars equal the number of residuals in that bar. The solid line represents the normal distribution and the dashed reference lines in the graph indicate the -3 to +3 data range where 99.7% of the data values would be expected to fall if they followed a normal distribution. In general, the residual distributions tended to have high central peaks with the majority of values near zero indicating that a large proportion of the predicted values were close to the observed data. The distributions for the rate and rate change models produced a few large, positive-valued residuals compared to one value >3 for the cumulative model. All three models failed the Shapiro-Wilk test for normality ($p < 0.001$, Table 2.3). This test can be very sensitive to departures from normality such as the high central peaks and although the distributions appeared to approximate a normal distribution, the distributions nevertheless failed the Shapiro-Wilk test. The QQ plots in Figure 2.6 show the departure from normality for the residuals, highlighting the heavy-tailed nature of the distributions.
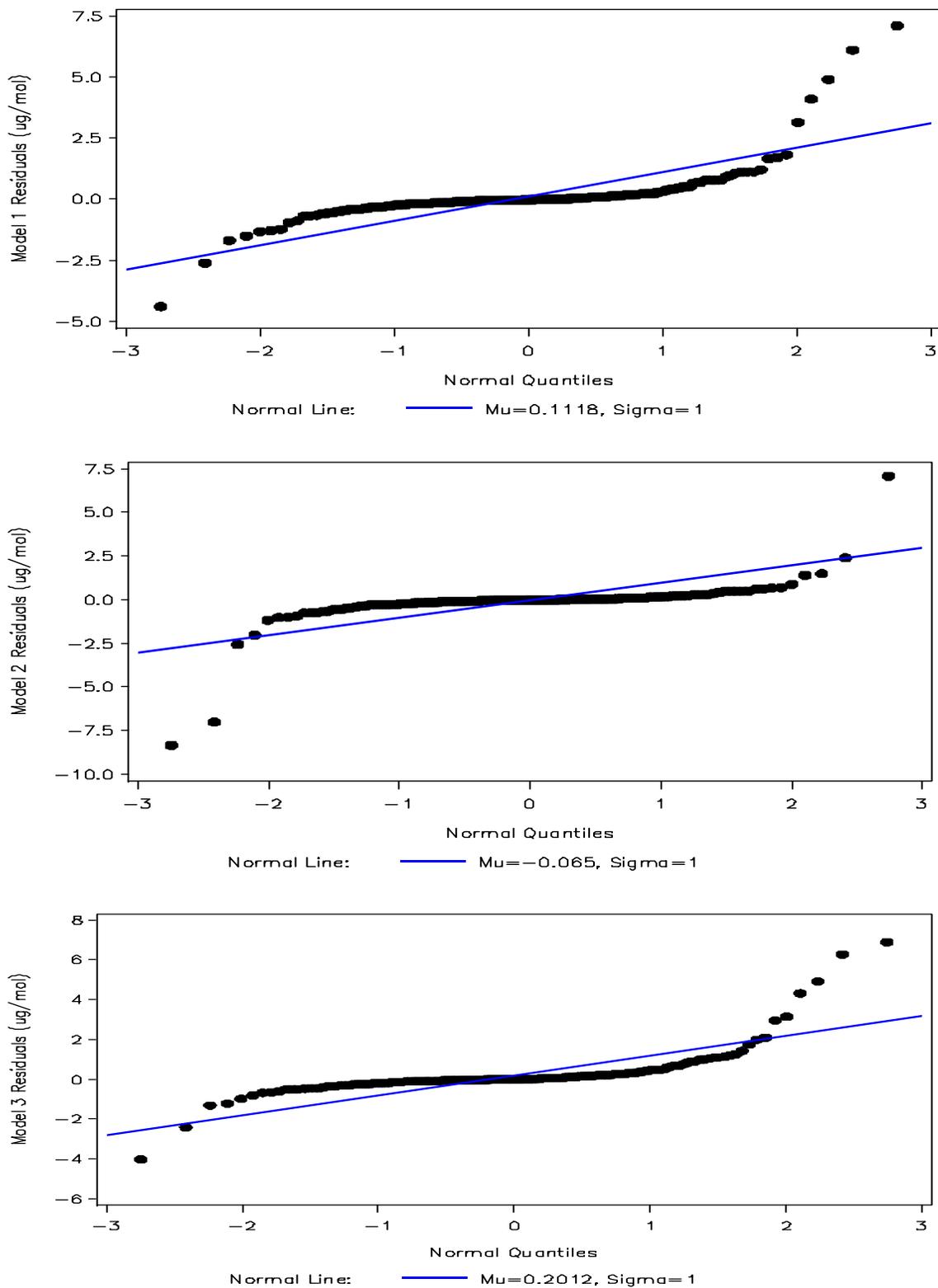
Figure 2.6: Quantile-Quantile Plots of Residuals

Location

    The Sign-rank test for location assessed whether the residual distribution from each of the models was centered on a value of zero. The mean of the Model 3 residuals was statistically significantly different from zero (p=0.0101), however, the null hypothesis for centering on zero was not rejected for Models 1 and 2 (p=0.9995 and p=0.2238, respectively). The mean standardized residual for Model 2 was closest to zero at -0.06 µg/mol versus 0.11 and 0.20 µg/mol for Models 1 and 3, respectively. Median values for all 3 models were close to zero with results showing medians of -0.015, -0.003, and 0.013 µg/mol for Models 1, 2, and 3, respectively. Analysis of the non-standardized residuals yielded identical results for the statistical tests; however, with means of 5.12, -8.95, and 11.17 µg/mol for Models 1, 2 and 3, respectively. Median non-standardized residual values for all 3 models were close to zero with values of -0.75, -0.48, and 0.71 µg/mol for Models 1, 2, and 3, respectively.

Table 2.3: Statistical Tests on Residuals

	Model 1 (N=205)	Model 2 (N=205)	Model 3 (N=205)
Shapiro-Wilk Test for Normality (p-value)	W=0.581 (p<0.001)	W=0.408 (p<0.001)	W=0.582 (p<0.001)
Signed-rank Test for Location (p-value)	S=0.5 (p=0.9995)	S=-1036.5 (p=0.2238)	S=2177.5 (p=0.0101)
Homogeneity of Variance **Levene's Test** **(p-value)**	L=1.03 (p=0.3808)	L=0.97 (p=0.4081)	L=1.34 (p=0.2618)
Brown-Forsythe Test **(p-value)**	F=1.54 (p=0.2060)	F=0.85 (p=0.4681)	F=2.14 (p=0.0967)

Residual values skewed away from zero as the 1-AP concentration increased for Models 1 and 3 (Figure 2.7) indicating an increasing bias with increasing 1-AP concentration. The dashed line in Figure 2.7 references zero, the expected center of the residual distribution. The solid line depicts the regression slope of the residuals, which is significantly different from zero for all three models (p<0.001).
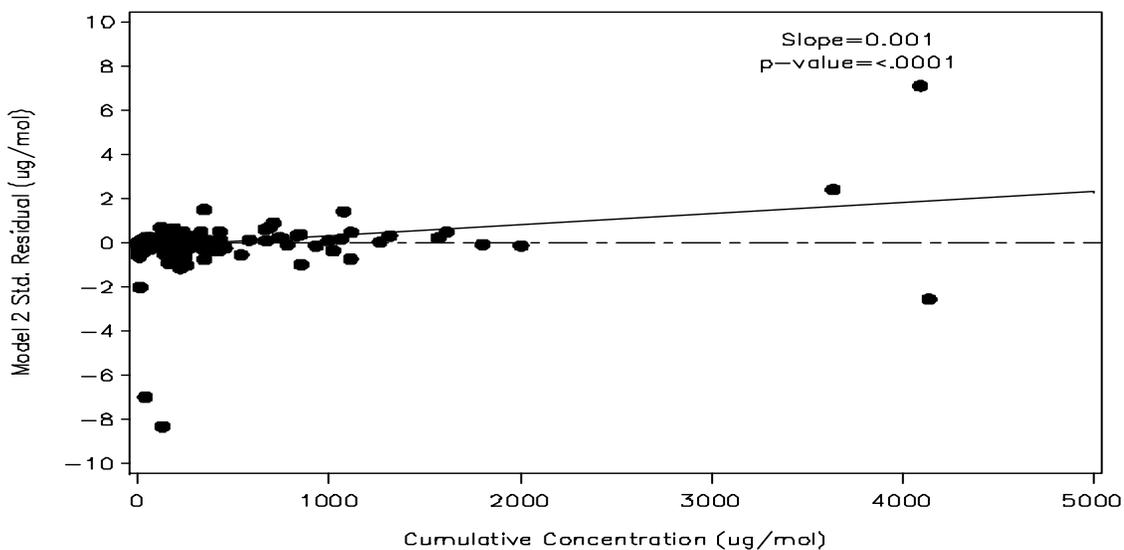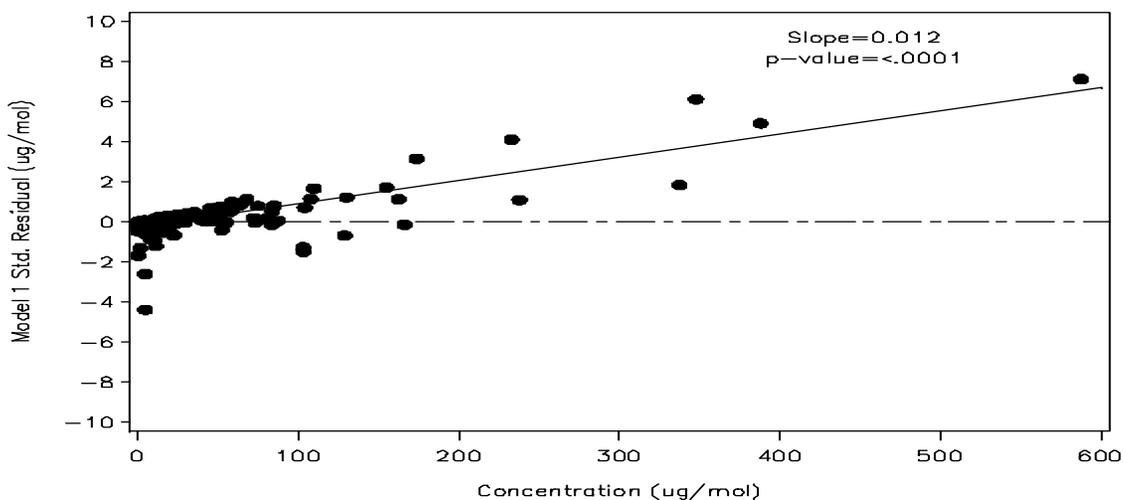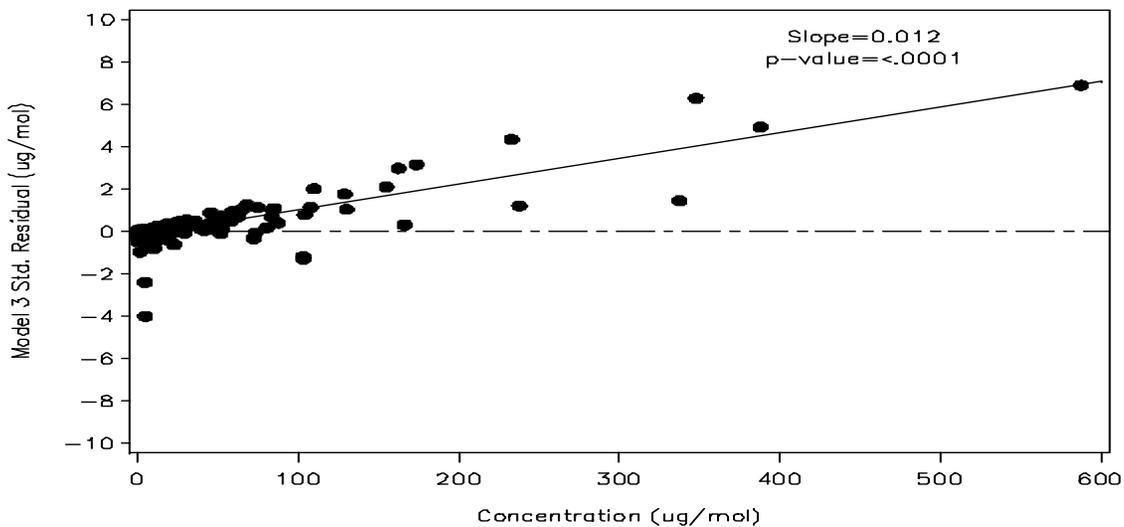
Figure 2.7: Standardized Residuals by 1-AP Concentration

Examination of the residuals over time showed a difference in patterns for Models 1 and 3 versus Model 2 (Figure 2.8). There was an increase in the number of large, positive-valued, residuals at the early time points for Models 1 and 3 compared to Model 2, indicating that Models 1 and 3 had a greater tendency to under-estimate the magnitude of the early peaks for some individuals. No clear pattern between standardized residuals and time was noted for Model 2.
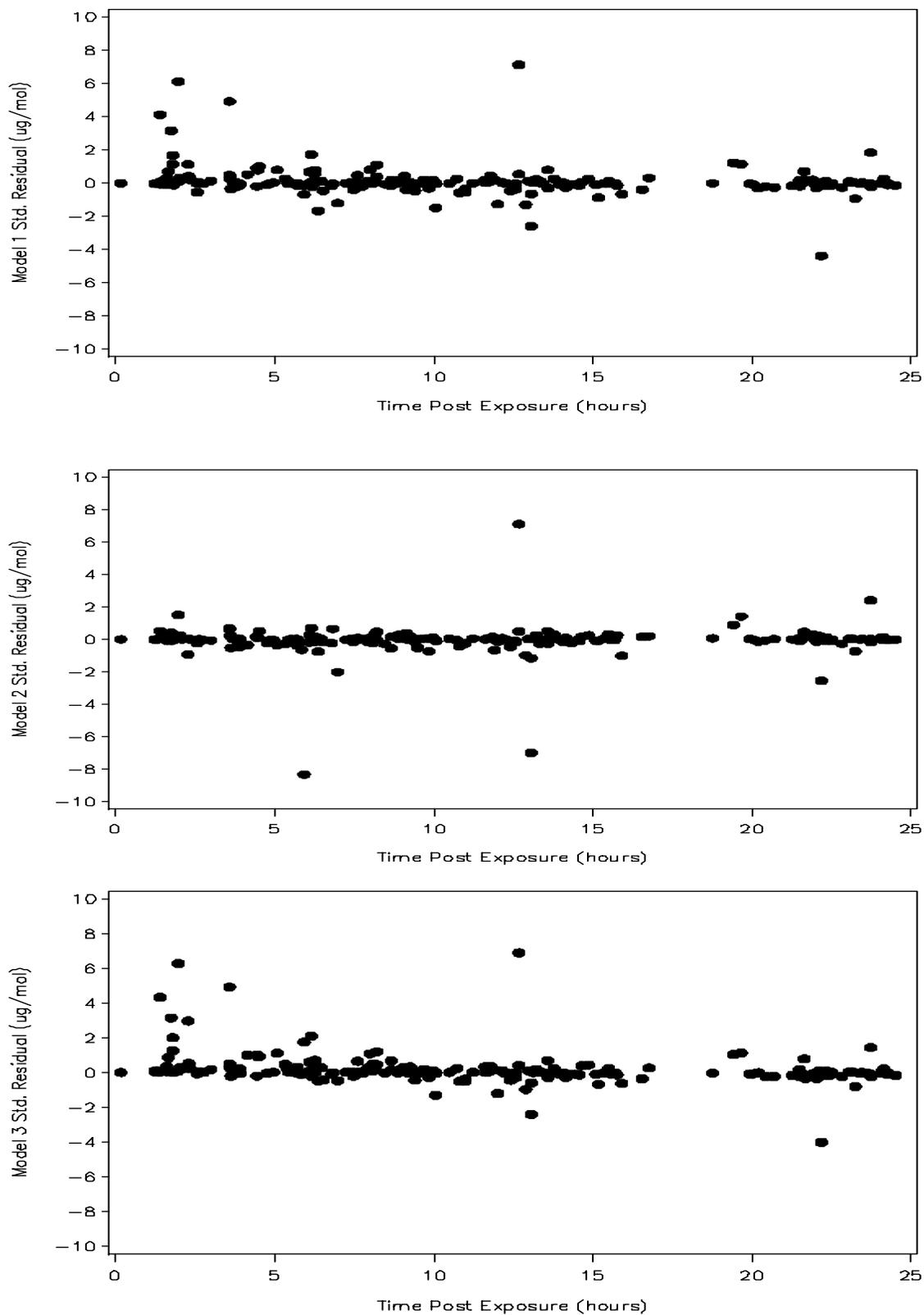
Figure 2.8: Standardized Residuals Over Time

Homogeneity of Variance

Tests for homogeneity of variance were conducted by first dividing time into four equal quartiles, such that each quartile contained approximately the same number of values. Samples were divided into quartiles defined by < 5.5 hours, 5.5-<9.85 hours, 9.85-15.55 hours and ≥15.55 hours as specified in Table 2.4 below.

Table 2.4: Time Quartiles

Quartile	Sample Collection Time (hours)	Number of Observations
1	< 5.50	51
2	5.50 – < 9.85	52
3	9.85 - < 15.55	50
4	≥ 15.55	52

Levene's and the Brown-Forsythe (BF) tests were conducted on each quartile comparing the variances of the residuals within each model across the quartiles. The BF results did not show any statistically significant differences in terms of variance across time within model 1 or 2 (p=0.21, 0.47), however, the BF results for model 3 showed a marginally statistically significant difference in variance across time (p=0.10). Using Levene's test, none of the Models had statistically significant differences in variance across time, p=0.38, 0.41, and 0.26 for Models 1, 2, and 3, respectively. One additional figure (Figure 2.9) based on the four time quartiles used for the homogeneity of variance assessment was created to review the standardized residual distribution within each quartile. The median value for each model and quartile group is provided in the figure. Review of the median residual value for each model shows that Model 2 is best at

providing a center value nearest to zero across the four time quartiles while Models 1 and 3 show a tendency to overestimate the early peaks of the first quartile and underestimate the late peaks in the fourth quartile, consistent with the data plotted in Figure 2.8.
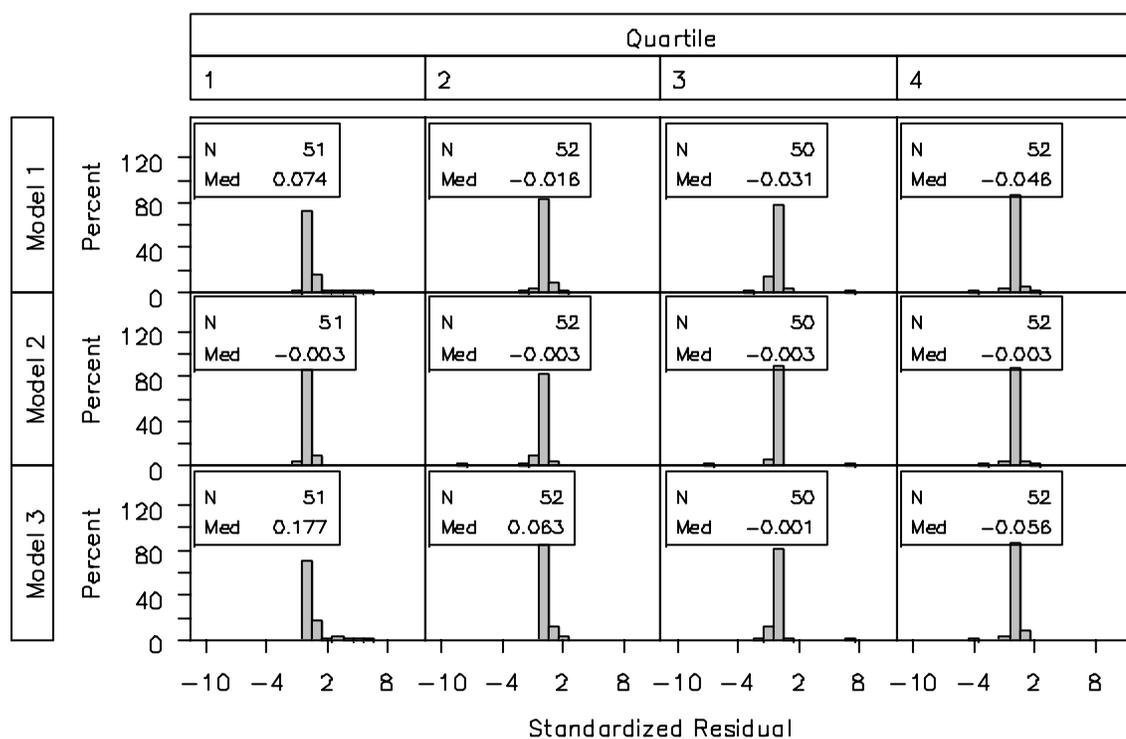
Figure 2.9: Standardized Residual Distribution by Time Quartiles

Sensitivity Analysis

An identical residual analysis conducted on all 55 DE exposed subjects was consistent with the results detailed above such that all 3 models failed the Shapiro-Wilk test for normality, however with consistent variability across time. Models 1 and 3 had a tendency to overestimate data at early time points and underestimate data at later time points.

Simulation

A simulated data set with more consistent subject urine excretion profiles was created to provide a second example of model fit comparisons. The purpose of the simulated data set evaluation was to assess whether the same model choice would be made when data was more consistent across subjects. The approach used to obtain additional consistency was to simulate concentration profiles based on parameters that would correlate with a limited range of maximum excretion times (Tmax). The method of data simulation and the selection of parameter values and their associated variances are detailed below.

The simulated data set consisted of 500 subjects with 4 to 8 urine samples per subject as in the evaluable subset of the DE study. The data were generated by first randomly determining the number of time points, n_i, for subject i. Let X_r $(r=1,...,5)$ denote the number of subjects with n_i urine samples where n_i was between 4 and 8, then the number of subjects with n_i urine samples followed a multinomial distribution where $X_r \sim Multinomial(0.2, 0.2, 0.2, 0.2, 0.2; 40)$. Spot sampling was mimicked by assuming collection times were randomly and uniformly distributed throughout specified periods according to the following distribution:

Sample Number	Time Value Simulation Model
1	Uniform(0, 2)
$>1 - \max(\text{int}(n_i/2))$	Uniform(2, 10)
$>\max(\text{int}(n_i/2)) - n_i-1$	Uniform(10, 20)
n_i	Uniform(20, 24)

Specifically, each simulated subject provided at least one sample from 0 to 2 hours and another from 20 to 24 hours; the remaining samples were split between the periods from 2 to 10 hours and from 10 to 20 hours post exposure.

Cumulative urine concentration values were generated in a two-step process. In the first step, PK parameter values were determined for each subject. The parameters followed a multivariate normal distribution specified as:

$$\begin{bmatrix} \log(K) \\ \log(Km) \\ \log(Ke) \\ \log(F) \end{bmatrix} \sim MVN \left(\begin{bmatrix} -1.91 \\ -1.42 \\ 1.22 \\ 2.40 \end{bmatrix}, \begin{bmatrix} 0.0365 & -0.022 & 0 & 0 \\ -0.022 & 0.020 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right)$$

where we note that parameters were exponentiated when modeled to assure positive estimates. In this parameter distribution, *Ke* and *F* were set as fixed values with no associated variance in order to minimize the variability of the simulated responses and obtain greater consistency in subject data as compared to the DE experiment. The expected values of log(*K*), log(*Km*), log(*Ke*), and log(*F*) were derived from the experimental data model fit by selecting subjects whose estimated time of maximum excretion (Tmax) was between 4 and 7 hours, and then finding the mean *K*, *Km*, *Ke*, and *F* for those subjects. The variance and covariance of log(*K*) and log(*Km*) were determined by setting the coefficient of variation (CV) to 10% and the correlation, $\rho_{K,Km}$, to -0.8 in order to minimize the variability of the simulated responses. In this manner, $\Sigma_{K,K} = (CV * \kappa)^2$ and $\Sigma_{Km,Km} = (CV * \kappa m)^2$ with $\Sigma_{K,Km} = \rho_{K,Km} * \sigma_K * \sigma_{Km}$; where κ and κm are the expected values of log(*K*) and log(*Km*), respectively; $\Sigma_{K,K}$ and $\Sigma_{Km,Km}$ are the random effects variances of log(*K*)and log(*Km*), respectively; and $\Sigma_{K,Km}$ is the random effects covariance of log(*K*) and log(*Km*). As with the experimental data, *D*, the amount of DE exposed to the subject, was not included in the modeling as this was standardized as part of the experiment.

In step two of the simulation process, y_{ij}, the cumulative urine excretion concentration for the ith subject ($i=1,\ldots,m$) at the jth time point, t, ($j=1,\ldots,n_i$) was calculated as

$$y_{ij} = \frac{FKeKmD}{Km-K}\left(\frac{1}{K}-\frac{1}{Km}\right)+\frac{FKeKmD}{Km-K}\left[\frac{1}{Km}e^{-Kmt}-\frac{1}{K}e^{-Kt}\right]+\varepsilon_{ij}$$

where the ε_{ij} are the within-subject errors and $\varepsilon_{ij} \sim N(0,0.5)$.

Analysis of the simulated data sets showed the number of subjects providing 4 through 8 samples in the simulated data set was similar to that for the DE experiment. Table 2.5 below shows the percent of subjects by the number of samples provided, this distribution is similar to that from the DE data shown in Table 2.2.

Table 2.5: Distribution of Number of Urine Samples in the Simulated Data Set

Number of Samples	Percent of Subjects	Cumulative Percent
4	25.2	25.2
5	21.8	47.0
6	27.6	74.6
7	17.8	92.4
8	7.6	100

Figure 2.10 below shows a sample distribution of the simulated spot-sampled urine concentration data for a subset of 40 subjects. This data is displayed as both cumulative values (left panel) and rates of excretion (right panel) with dots representing the simulated values and lines drawn to connect the values and show trends. The simulated data tended to showed more variety in terms of response patterns when

viewing the rate values. This may be due to the sparseness of the "spot" collections in the simulated data.
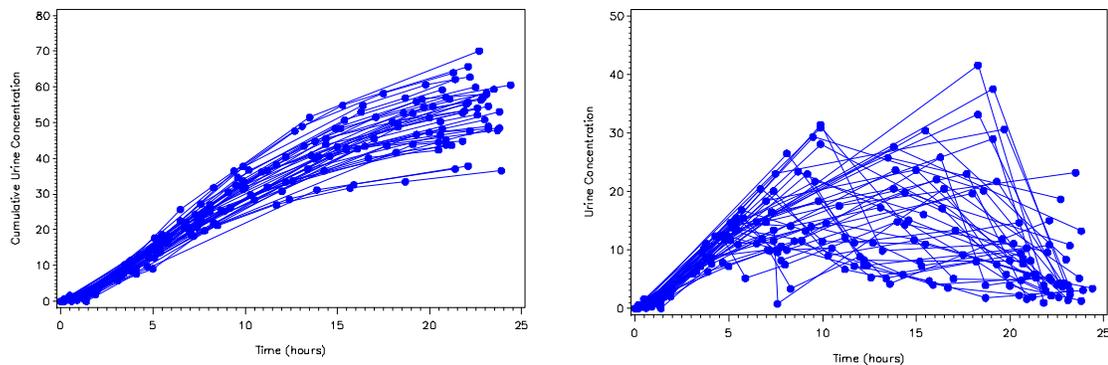
Figure 2.10: Simulated Urine Concentration Data – Left Panel shows cumulative values, Right Panel shows excretion rate values

The Tmax values of the 500 simulated subjects ranged from 4.4 to 6.2 hours with a median of 5.2 hours. These values were similar to the experimental data upon which the simulation parameters were based. The simulated data was evaluated using the same three models used for the DE data, the rate model (Model 1), the cumulative concentration model (Model 2) and the change in cumulative concentration model (Model 3) as previously detailed in the Statistical Methods section. Standardized residuals were compared graphically and with statistical tests for normality, location, and homogeneity of variance across time.

Normality

Histograms comparing the distributions of the standardized residuals from the three models showed that most values fell between -3 and +3 standard deviations and distribution curves were approximately bell-shaped, albeit with high central peaks for Models 1 and 3 (Figure 2.11). Mean and median statistics for Model 2 were closest to zero while Models 1 and 3 had several values which were more than 3 standard deviations above the mean. All 3 models failed the Kolmogorov-Smirnov test for normality KS=0.08, 0.02, 0.06 for Models 1, 2, and 3, respectively with $p < 0.01$ for all 3 tests). Note that due to the increased sample size of the simulated data set, the Kolmogorov-Smirnov test was used in place of the Shapiro-Wilk test for testing the normality assumption.
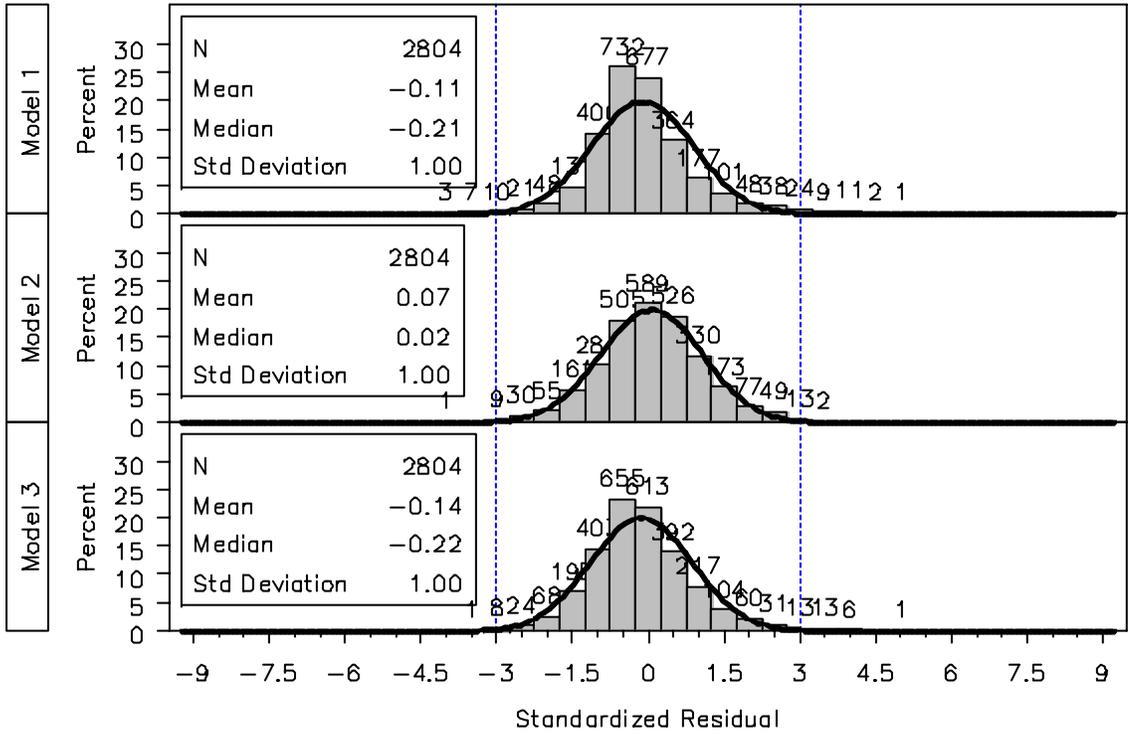
Figure 2.11: Simulated Data: Standardized Residual Distribution

QQ plots of the standardized residuals show that the distributions followed a normal distribution for Model 2 while the right skew is noticeable in the QQ plots for Models 1 and 3 (Figure 2.12).
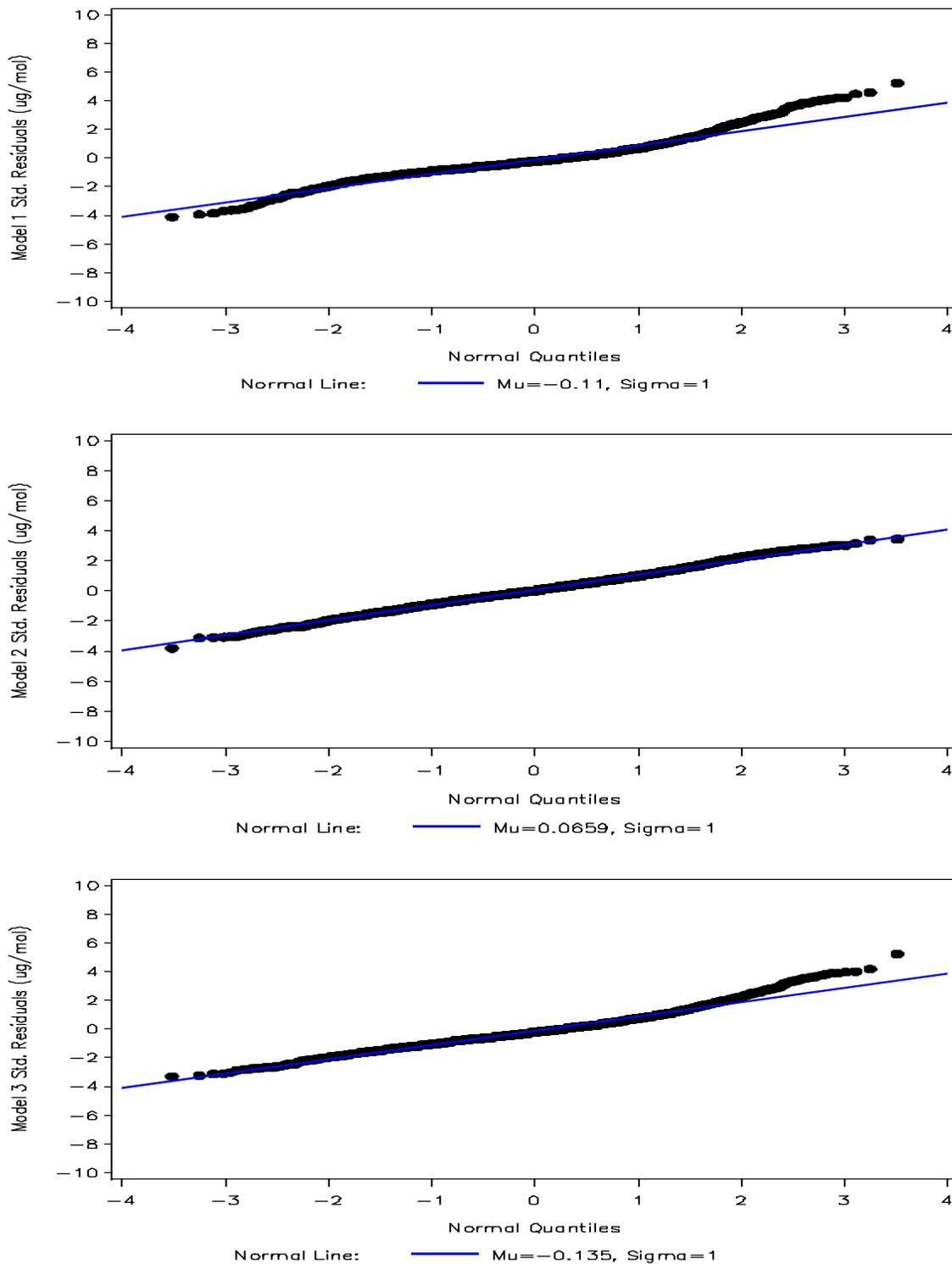
Figure 2.12: Simulated Data: QQ Plots of Standardized Residuals

Location

The mean (median) standardized residuals for Models 1, 2, and 3 were -0.11 (-0.21), 0.07 (0.02), and -0.14 (-0.22) μg/mol. The Sign Rank test for all three models was statistically significant (p<0.005), indicating that the means of the standardized residuals were different from zero. Examination of the standardized residuals by concentration and over time are displayed in Figures 2.13 and 2.14. Reviewing the standardized residuals by concentration (Figure 2.13), a clear trend for increasing residuals with higher concentrations was noted for Models 1 and 3. No pattern for residuals with changing concentrations was seen for Model 2.
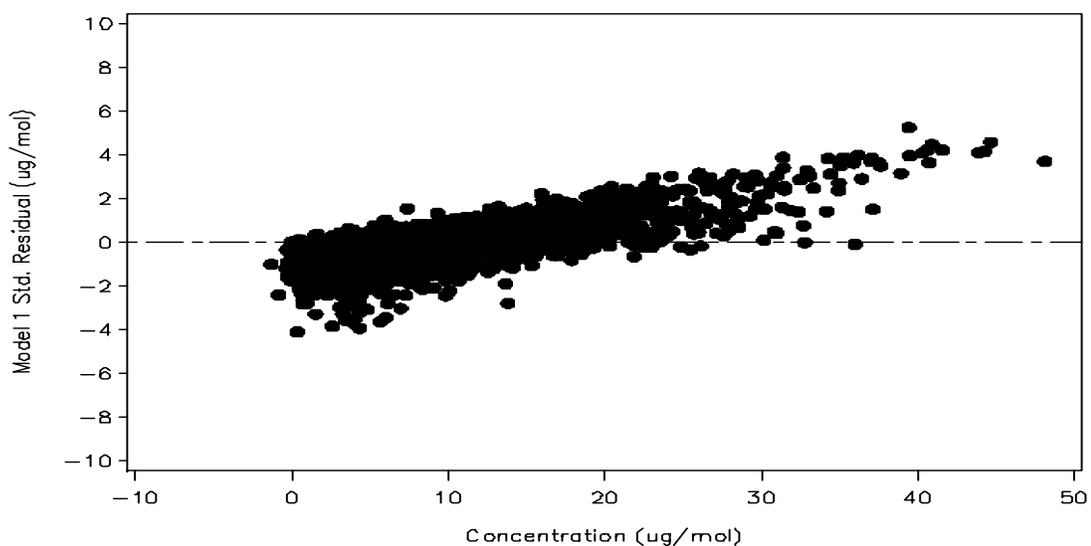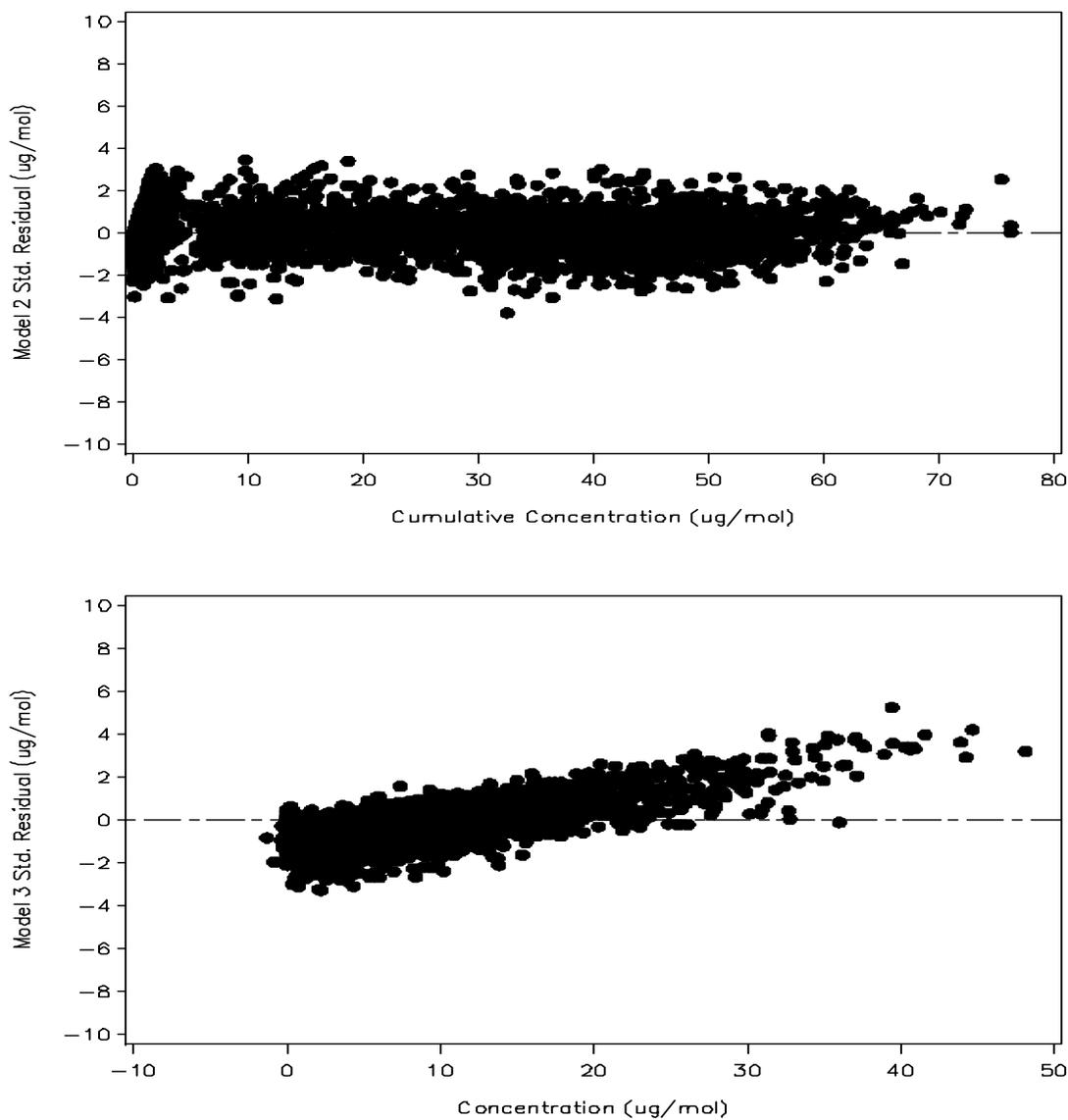
Figure 2.13: Simulated Data: Standardized Residuals by Concentration

No time trends were apparent for Model 2 based on the graphs displayed in Figure 2.14. For Models 1 and 3, residuals at the early and late time points tended to have negative values indicating the predicted value may be overestimating the data at these time points. Ideally, residuals would be equally balanced above and below zero across all time points.
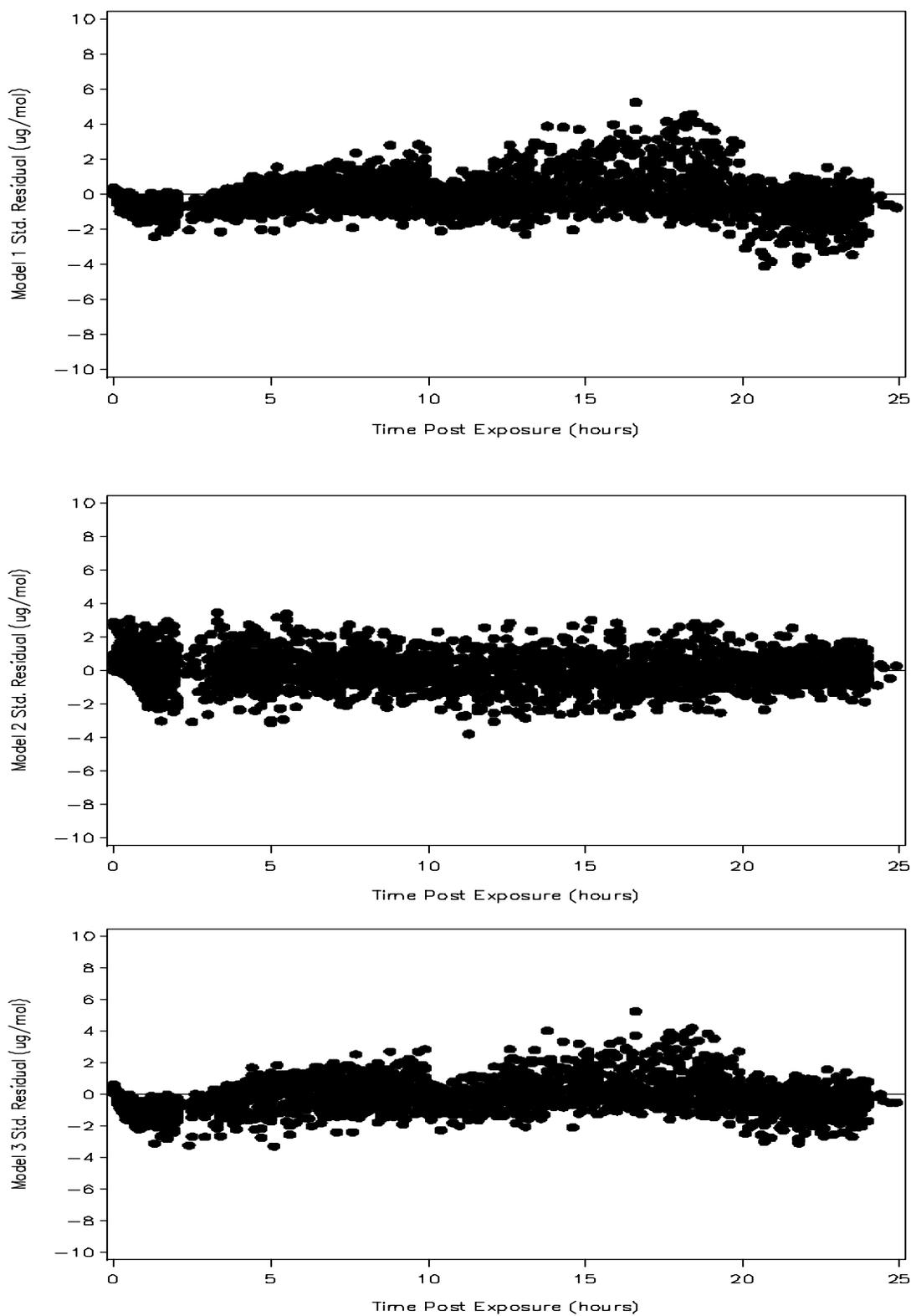
Figure 2.14: Simulated Data: Standardized Residuals Over Time

Homogeneity of Variance

The simulated data was divided into four time periods in order to examine the variances of the residuals within each model across time. Time periods of 0 - < 4.5 hours, 4.5 – <10.1 hours, 10.1 – < 18.1 hours, ≥18.1 hours were used to divide the standardized residuals into approximately equal quartiles. Levene's and the BF test were then used to assess the homogeneity of variance across time within each model. The results of the BF showed that the variation in residuals differed across time for all Models, ($p<0.01$). Similar results were obtained with Levene's test. Figure 2.15 below displays histograms of the standardized residuals within each time period and model.
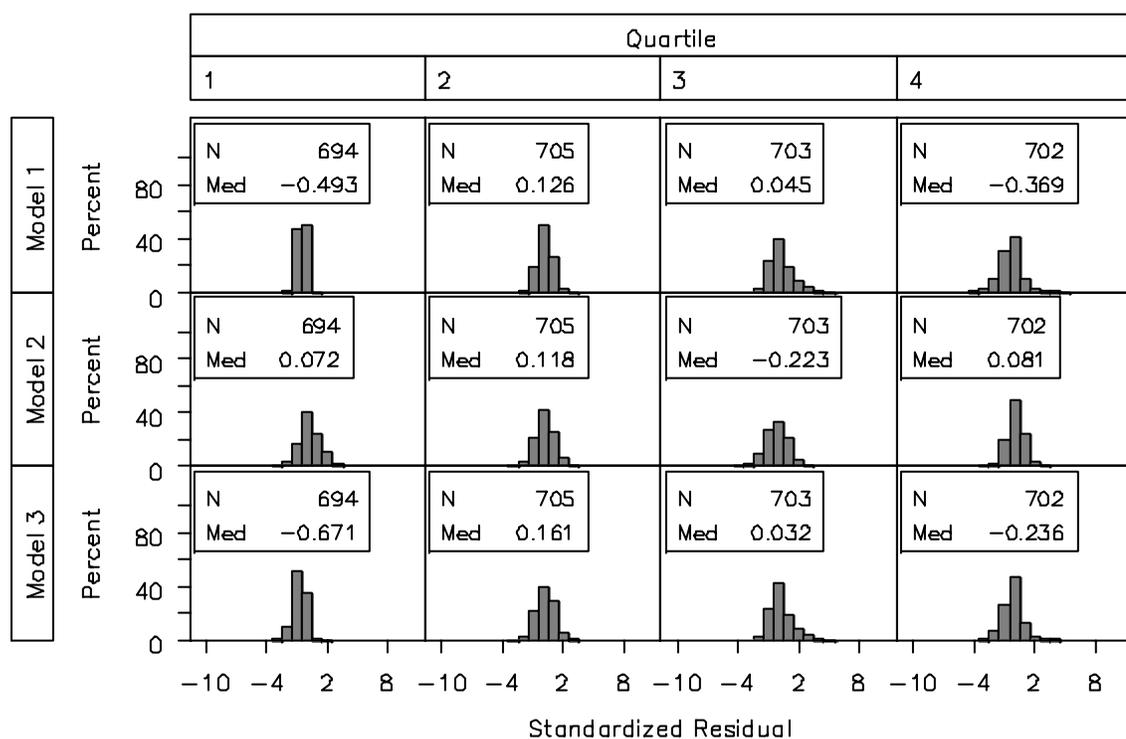
Figure 2.15: Simulated Data: Standardized Residuals by Time Quartiles

Discussion

In many nonlinear modeling applications, residual values are assumed to (1) follow a normal distribution; (2) be centered on zero; and (3) possess a variance that is homogenous across the model parameters. Comparisons of three models were conducted using data from a DE exposure experiment and on simulated data. Standardized residuals were analyzed and graphed to evaluate normality, location, and homogeneity of variance.

Results using the DE data indicated that none of the models produced residuals that were viewed as following a normal distribution according to the Shapiro-Wilk test. Visual review revealed that the majority of the values tended to follow a normal distribution with high central peaks although all models had a small number of large, mostly positive-valued residuals. The Model 3 residual median was found to be statistically significantly different from zero, a violation of the second assumption listed above, which indicated a potential bias in the predicted values and parameter estimates. The variance for all three sets of residuals was consistent across time.

Visual examination of the residuals across the 1-AP concentrations and collection times revealed that higher 1-AP concentrations were associated with higher residual values for Models 1 and 3, indicating a positive correlation where there should be none. Model 2 also had some large-valued residuals at higher 1-AP concentrations; however, no determination of bias could be made as these residuals were more sparse and evenly distributed around the expected value of zero compared to the residuals of Models 1 and 3. Examination of the residuals over time revealed that early time points tended to be underestimated for Models 1 and 3, but not for Model 2, indicating that Models 1 and 3 may not be providing as good a fit as Model 2.

A simulated data set was constructed to ascertain the model selection process in the case of more consistent urine concentration data. Results assessing normality were similar between the three models. As with the DE data, all three models failed the test for normality, although visually, all residual distributions generally followed bell-shaped curves and Model 2 appeared to better follow a normal distribution according to the QQ plot. Location tests revealed that the median residual values were statistically different from zero for all models, although the mean and median for Model 2 were closest to zero among the three models. Trend evaluations by concentration showed patterns of response existed for Models 1 and 3 indicating some potential systematic bias for these models, and no response pattern for the cumulative excretion model, Model 2. In addition, all Models showed heterogeneous variance across time quartile according to Levene's test and the Brown-Forsythe test.

Based on the results of the statistical tests and the visual examination of the residuals, Model 2 was deemed optimal in terms of goodness-of-fit compared to Models 1 and 3. In both the DE and simulated data, concentration trends were noted for Models 1 and 3 and not with Model 2. The cumulative data model, Model 2, should be utilized when modeling this DE urine excretion data based on the observed data analysis and should be considered when future work calls for modeling urine excretion data in general.

Significant challenges existed in the modeling of the DE urine excretion data. A large variety of subject response profiles made population-level parameter estimation unachievable. Possible contributors to the variations in response included differences in sampling times, numbers of samples per subject, potential variations in competing environmental exposures prior to and after the subjects were in the clinic as well as the

loss of some samples due to out-of-range creatinine measures (required for standardization). Another important potential source of variability is the individual's response to DE exposure due to differences in 1-NP metabolism between subjects. Despite these challenges, it was important to identify a model that best fits the data in order to obtain parameter estimates that describe the data even when fit on an individual subject-level basis. Future work will examine the individual subject-level parameter estimates for calculation of time of maximum excretion (Tmax) as well as potential grouping of subjects according to response profiles to ascertain differences.

Spot urine sampling is an important public health tool for biomonitoring purposes. Analyses of the time and concentration profiles utilize pharmacokinetic techniques based on nonlinear models. Residual analysis to assess goodness-of-fit for model comparisons is an important part of the model selection process. Modeling the cumulative DE urine excretion values smoothed out differences in the response profiles over time and this may have resulted in the ability of the cumulative model to better predict the actual data. By looking at the data in different ways, e.g. cumulative values vs. rate of excretion, we were able to identify a better fitting model than if we had limited ourselves to just one approach. The evaluation of multiple models for goodness-of-fit should become part of our standard practice when assessing nonlinear data.

References

Bamford, H.; Baker, J. Nitro-polycyclic aromatic hydrocarbon concentrations and sources in urban and suburban atmospheres of the Mid-Atlantic region. *Atmos. Environ. 2003. 37(15),* 2077-2091.

Barr, D.B.; Wang, R.Y.; Needham,L.L. Biologic Monitoring of Exposure to Environmental Chemicals throughout the Life Stages: Requirements and Issues for consideration for the National Children's Study. *Environ. Health Perspect.* 2005. *113(8),* 1083-1091.

Beland, F. A.; Kadlubar, F. F. Metabolic Activation and DNA Adducts of Aromatic Amines and Nitroaromatic Hydrocarbons. In *Carcinogenesis and Mutagenesis I* ; Cooper, C.S.; Grover, P.L., Eds.; Springer-Verlag: New York, 1990, 267-325.

Bond, J. A., Sun, J. D., Medinsky, M. A., Jones, R. K. , Yeh, H. C. (1986). Deposition, Metabolism, and Excretion of 1-[^{14}C]Nitropyrene and 1-[^{14}C]Nitropyrene Coated on Diesel Exhaust Particles as Influenced by Exposure Concentration. *Toxicology and Applied Pharmacology.* 85, 102-117.

Dutcher, J. S., Sun, J. D., Bechtold, W. E., Unkefer, C. J. (1985). Excretion and Metabolism of 1-Nitropyrene in Rats after Oral or Intraperitoneal Administration. *Toxicological Sciences*, 5 (2), 287-296.

Howard, P. C., Consolo, M. C., Dooley, K. L., Beland, F. A. (1995). Metabolism of 1-Nitropyrene in Mice: Transport Across the Placenta and Mammary Tissues. *Chemico-Biological Interactions,* 95, 309-325.

IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. http://monographs.iarc.fr/ENG/Monographs/vol46/volume46.pdf (accessed July 30, 2008). Volume 46 Diesel and Gasoline Engine Exhausts and Some Nitroarenes.

Laumbach, R. (2009). Quantification of 1-aminopyrene in human urine after a controlled exposure to diesel exhaust. *Journal of Environmental Monitoring*, (11) 153-159.

SAS Institute Inc. *SAS OnlineDoc®,* Version 9.1. SAS Institute Inc.: Cary, NC, 2004.

Seidel, A.; Dahmann, D.; Krekeler, H.; Jacob, J. Biomonitoring of polycyclic aromatic compounds in the urine of mining workers occupationally exposed to diesel exhaust. *Int. J. Hyg. Environ. Health.* 2002. *204,* 333-338.

Shargel, L. & Yu, A. B. C. (1993). *Applied Biopharmaceutics and Pharmacokinetics.* Stamford: Appleton & Lang.

S-Plus 6.0 Professional Release 2 for Windows

Sun, J. D., Wolff, R. K. , Aberman, H. M. , McClellan, R. O. (1983). Inhalation of 1-nitropyrene Associated with Ultrafine Insoluble Particles or as a Pure Aerosol: A Comparison fo Deposition and Biological Fate. *Toxicology Applied Pharmacology,* 69 (2) 185-198.

Tokiwa, H.; Ohnishi, Y. Mutagenicity and carcinogenicity of nitroarenes and their sources in the environment. *Crit Rev Toxicol.* 1986. *17(1),* 23-60.

Chapter 3

Patterns in Times to Maximum Urine Excretion of 1-aminopyrene Following Diesel Exhaust Exposure

Abstract

1-Nitropyrene is the most abundant nitro-PAH, a class of carcinogens identified in diesel exhaust particles. Urine excreted 1-aminopyrine (1-AP), a metabolite of 1-nitropyrene, has been suggested as a biomarker for diesel exhaust exposure. In our recent study, a series of spot urine samples were collected in a group of healthy volunteers following a controlled diesel exhaust exposure. These urine samples, collected within 24 hours after the exposure, were analyzed for 1-AP. This chapter focuses on characterizing the urine excretion profile and estimating the time of maximum excretion. A one-compartment first-order pharmacokinetic model was fit using nonlinear modeling techniques. Summarization of the time of maximum excretion proved difficult due to the variety of subject response profiles. Simulations were used to help understand the observed excretion profiles and generated a hypothesis that there may be at least two sub-populations with different response profiles present. Results showed that 70% of the subjects had a median time of maximum excretion of 5.9 hours, while 30% of the subjects may have had maximum excretion times longer than 24 hours.

Introduction

Pharmacokinetic models are used to describe the time and concentration profiles of drugs and other chemicals following exposure. These models can be used to estimate

pharmacokinetic parameters such as absorption and excretion rates. While the typical pharmacokinetic analysis is based on data from plasma samples, modeling of urine samples represents an important arena. Urine collection is less invasive than plasma sampling and does not require subject presence at the clinic site at the time of sample collection as voids can be collected and retained under appropriate storage conditions for later delivery to the clinic site or laboratories.

Urine biomonitoring may be used in cross-sectional or cohort studies of environmental-health associations in large populations, for example, exposure to diesel exhaust or diesel exhaust particles. Thus, characterization of the trends in concentration profiles in urine may be important for understanding results and for improving and simplifying studies, as urine collection can be made at an optimal time to capture the exposure. Once preliminary experiments have established a range of values for the time of maximum excretion (Tmax), collection of spot urine samples may be a viable strategy for assessing exposure. In addition, simplification of the collection and timing of samples would bring efficiency to the biomonitoring process, reducing costs and encouraging subject compliance by decreasing the number of samples required.

In a recent experiment, spot urine samples representing partial voids were collected from healthy volunteers during a 24-hour period following a brief controlled exposure to diluted diesel exhaust (DE). In the experiment, each study participant underwent two one-hour exposure sessions: one was a diluted DE atmosphere and the other was a filtered clean air atmosphere (CA). The overall goal of the experiment was to examine whether 1-aminopyrene (1-AP) could serve as an adequate biomarker for DE exposure. Published results indicate a significant difference between DE and CA

sessions with respect to the average urine concentrations of 1-AP in the 24 hours following exposure [Laumbach et al., 2009]. An additional goal was to characterize the excretion time course of the biomarker 1-AP, in order to optimize sampling times in future studies.

This study was undertaken because exposure to diesel exhaust constitutes a public health concern as DE contains a large suite of toxic chemicals including nitro-polycyclic aromatic hydrocarbons (nitro-PAHs) [Bond et al., 1986, Howard et al., 1995]. Many nitro-PAHs (e.g., nitropyrenes) have been shown to have genotoxic and carcinogenic properties in experimental animals [Beland and Kadlubar, 1990, IARC Monogrpahs, Tokiwa and Ohnishi, 1986]. Concerns about the public health risks resulting from inhalation exposures to nitropyrenes have inspired numerous studies to further the scientific knowledge on the deposition, metabolism and excretion of nitropyrenes [Bond et al., 1986, Butcher et al., 1985, Sun et al., 1983]. Results of these studies have shown that nitro-PAHs are rapidly absorbed and metabolized. A recent study found that concentrations of nitro-PAHs were 3 orders of magnitude higher in DE particles than in urban ambient particles, further supporting the notion that nitro-PAHs are highly specific to diesel combustion emissions [Bamford and Baker, 2003]. The most abundant nitro-PAH in DE particles was 1-nitropyrene (1-NP). A metabolite of 1-NP, 1-aminopyrene (1-AP) has been measured in the urine of DE-exposed miners [Seidel et al., 2002]. However, little is known about the detailed timing of 1-AP excretion following a spiked DE exposure; and the present chapter attempts to address this specific question.

Visual review of the 1-AP urine excretion profiles following DE exposure in the current study revealed a wide variety of responses among subjects. As a result of these

varied responses, no single set of pharmacokinetic parameter values was identified which would allow for fitting a model of the 1-AP urine excretion profiles for the whole group of study participants together. Therefore, the main goal of the present chapter is to explore how the distribution of estimated peak excretion times is impacted by differences in subject absorption and excretion rates as well as sampling schemes in order to better understand the 1-AP urine excretion rates observed in the DE-exposed subjects. First, the observed experimental DE data was analyzed for each subject in order to provide a realistic distribution of excretion times and for comparison to simulated data. The simulated data representing spot sampling and fixed sampling times were examined using urine concentration profiles with early peaks (1-2 hours), mid-range peaks (3-4 hours) and late peaks (8-9 hours). Low and high parameter variability were also explored. Additional data was simulated representing a mixture of sub-populations of subjects in order to characterize a possible population panel that would be even more consistent with the observed results.

Materials and Methods

Diesel Exhaust Experiment.

In a recent study of DE exposure using a two-period crossover design, subjects underwent one hour of controlled exposure to either filtered clean air ("CA") or diluted DE (300 μg/m3 as $PM_{2.5}$) for one hour on separate mornings at least one week apart in a randomized order. Biomarkers were measured in 55 healthy volunteers using spot urine samples collected over a 24-hour period following exposure. Subjects were asked to

collect all urine voids (or as many as they could) within the 24 hours. Details of the experiment are provided in a previous publication [Laumbach et al., 2009].

In this chapter, we focus primarily on samples taken during the DE session. Since this study followed a 2-period crossover design, the DE exposure results were generated in both periods, with approximately half the subjects being exposed to DE in the first period and the other half being exposed in the second period. A partial void, spot sampling method was used to collect urine samples from study participants; hence, measurements of voided urine volume were not available. As concentration is a function of rate of excretion, volume voided and time, methods for standardization of biomarker concentration have been developed in order to accommodate partial-void sampling, generally used for biomonitoring purposes. The routine method used for this standardization is an adjustment based on urinary creatinine concentration where the biomarker is reported as a mass per unit of creatinine. This works well when making comparisons within individuals because the intra-individual variation in amount of creatinine excreted per day is relatively low [Barr et al., 2005]. Creatinine-standardized concentrations of urine biomarkers have been calculated for the data in this study so that comparisons within and between subjects could be made that would reflect the amount of 1-AP excreted accounting for differences in volume of output.

Approximately 64 of the 372 (17%) total DE and 51 of 356 (14%) total CA urine samples were considered invalid due to creatinine levels that were either missing, too low (<30 mg/dL) or too high (>300 mg/dL). This led to the development of rules for inclusion of a subject's data for model evaluation. These rules were similar to those used in a previous publication of this study [Laumbach, et al., 2009. Specifically, to be

included in the analysis data set, a subject must have provided at least two valid urine samples post exposure and have had no more than one invalid sample in a row, which resulted in a final evaluable set of 40 subjects. A sensitivity analysis which included all subjects and all samples was conducted to assess the impact of the missing data on the results.

Data Analysis

Prior to modeling the concentration profiles following DE exposure, an ANOVA model for crossover experiments was used to ascertain whether there were any statistically significant effects due to period or sequence. The response variable used to evaluate sequence and period effects was the time-weighted average concentration calculated as

$$TW_{hi} = \sum_{j=2}^{n_i} y_{hij}^* \left(t_{hij} - t_{hij-1} \right) \Bigg/ \sum_{j=2}^{n_i} \left(t_{hij} - t_{hij-1} \right) \tag{3.1}$$

where TW_{hi} is the time-weighted average for subject i following exposure to DE or CA, and h=DE or CA. Cumulative urine concentration values were calculated for the observed data. Specifically, if y_{ij}^* is standardized value for the ith subject (i=1,…, m) at the jth time point, (j=1,…, n_i) then $y_{ij} = \sum_{j=2}^{n_i} \left(t_{ij} - t_{i,j-1} \right) \times y_{ij}^*$ is the cumulative concentration for subject i at time j where j are the post-exposure time points and $y_{ij}=y^*_{il}$ at j=1. The trends in these cumulative time points were characterized using standard PK models. A previous analysis of the observed DE data showed that modeling the cumulative urine concentrations values provided a better fit compared to modeling the rate of urine excretion, so cumulative concentration values were used [Chapter 2].

Specifically, time of maximum excretion values were determined using estimated parameters derived from fitting a modified one-compartment first-order PK model for total urine excretion given by Equation 3.2:

$$D_u(t) = \frac{FKeKmD}{Km-K}\left(\frac{1}{K}-\frac{1}{Km}\right) + \frac{FKeKmD}{Km-K}\left[\frac{1}{Km}e^{-Kmt} - \frac{1}{K}e^{-Kt}\right] \qquad (3.2)$$

where $\dfrac{FkeKmD}{Km-K}\left(\dfrac{1}{K}-\dfrac{1}{Km}\right)$ and $\dfrac{FKeKmD}{Km-K}$ are constants with F representing the fraction of 1-NP absorbed, Ke the first order renal excretion constant, Km the first order rate constant for metabolism, K the first order rate constant for elimination, and D the amount of DE exposed to the subject [Shargel and Yu, 1993] . The model modification arose from switching Ka, the first-order absorption rate constant to Km, the rate of change from parent to metabolite, in recognition of modeling 1-AP, a metabolite of the parent compound, 1-Nitropyrene. Note that because subjects were exposed to the same DE concentration as determined by the experimental design, D was indistinguishable from F. Using this first-order rate model, the maximum likelihood equation for time of maximum excretion may be calculated using Equation 3.3:

$$Tm = \frac{\log(Km/K)}{(Km-K)} \qquad (3.3)$$

Due to the extent of variations in 1-AP urine excretion rates among subjects in the observed study data, no single set of parameter values was identified that applied to the entire evaluable DE-exposed subject population as a whole. Thus, the urine profiles were analyzed for each subject separately with estimation of fixed effects and within-subject variance only. Specifically, assuming $E(y_{ij})=Du(t_{ij})$, for each subject individually, the creatinine-standardized concentrations of 1-AP were modeled using SAS Proc NLMixed

[SAS Institute Inc., 2004] for nonlinear mixed effects to fit a model based on the cumulative concentration over time. Based on the results of previous work for model selection, the Laplace method was used to approximate parameter estimates for individual subjects. Parameters estimated from nonlinear modeling were used to calculate Tmax for each subject using Equation 3.3. For expositional purposes, we define the following notation:

Tmax = True time of maximum excretion (used with simulated data);

To = Observed collection time for the sample with maximum urine concentration; and

Tm = Modeled time of maximum urine excretion.

The subscript "p" will be added to denote a population-level value and the subscript "i" will be added when reference is made to individual subject-level values, e.g. Tm_i will denote the subject-level time of maximum excretion calculated from nonlinear model parameter estimates. The 1-AP To_i and Tm_i were summarized and compared using median, minimum and maximum values. Histograms were created of the void times, To_i and Tm_i.

Data Simulations

Tirteen types of simulated data sets were created to explore the impact of three separate factors on the derivation of To_i and Tm_i: sampling scheme, true theoretical peak time, and parameter variability. Factor levels included: (a) two levels of sampling schemes - fixed (preset) and spot (random) sampling; (b) three levels of true theoretical peak time - early (1 to 2 hours), mid- (3 to 4 hours) and late (8 to 9 hours); and, (c) two

levels of parameter variability - low or high. Full details of the data simulation process are provided in Appendix A. In general, values for K_i and Km_i were determined from the pre-specified **Tmax$_p$** and pre-specified level of parameter variability. Using randomly generated time points based on the sampling scheme, cumulative urine excretion values were simulated using Equation 3.2. **Tmax$_i$** were calculated from K_i and Km_i using Equation 3.3. The simulated data was analyzed using the same nonlinear modeling techniques as for the experimental data. *To$_i$* were derived directly from the simulated values and *Tm$_i$* were calculated using the nonlinear model parameter estimates. Summary statistics, including mean, minimum and maximum values, were used to compare **Tmax$_i$**, *To$_i$* and *Tm$_i$* results. Histograms were created of the simulated void times, **Tmax$_i$**, *To$_i$* and *Tm$_i$*.

Results

Experimental Data

The 40 subjects, whose data were used in the present data analysis, consisted of 25 males and 15 females with an average age of 24.0 years (range: 19-43 years). Eighty-five per cent of these subjects provided 5 or more valid post-DE-exposure urine samples with the majority (70%) of subjects providing between 5 and 7 valid samples, details are provided in Table 3.1.

Table 3.1. **Number of Valid Post Diesel Exhaust Exposure Urine Samples in the Evaluable Data Set**

Number of Samples	Number of Subjects	Percent

Number of Samples	Number of Subjects	Percent
3	1	2.5
4	5	12.5
5	9	22.5
6	11	27.5
7	8	20.0
8	6	15.0

Results of the ANOVA for the crossover design showed that there were no statistically significant period or sequence effects, indicating that the DE exposure data could be pooled into one analysis set (p=0.19 and 0.31, respectively).

Twelve (30%) of the 40 subjects had atypical urinary excretion rate curves in that the highest rate of excretion occurred at the end of the 24-hour period. This type of response pattern could have resulted from a secondary exposure or may have indicated that To_i had yet to be attained and may have occurred later than 24 hours after exposure. Figure 3.1 below illustrates this response profile using data from one example subject. For these subjects, Tm_i were extremely large and not interpretable. In such cases, Tm_i results were arbitrarily capped at 30 hours to simplify reporting. (The samples were collected up to 24 hours only.)

Figure 3.1. 1-AP concentration (μg/mol) over time following DE exposure for a representative subject demonstrating a late peak.
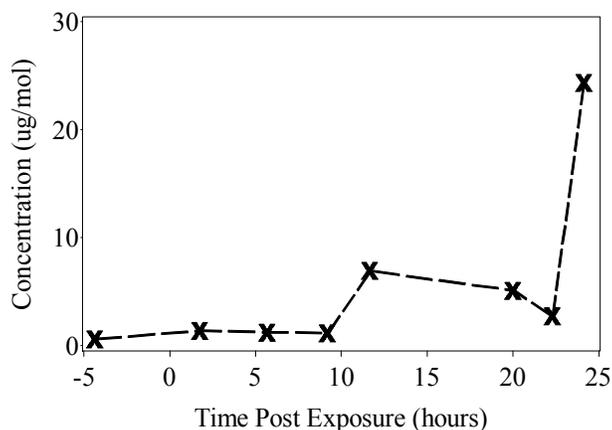
A comparative histogram of the experimental DE-exposure data for excretion of 1-AP showing the void times (row 1), To_i (row 2) and Tm_i (row 3) is presented in Figure 3.2. A total of 205 samples from 40 subjects were provided in the evaluable data set. A nearly flat distribution of void time data was observed and sampling times ranged throughout the collection period from 0.2 to 24.5 hours. Slightly fewer voids were collected during the 1.5 hour period immediately following exposure and at approximately 18 hours following exposure when it is likely that most study participants were sleeping. Although the span of values for To_i was nearly as wide as the void times, i.e. from 1.4 to 24.2 hours, the To_i appear as two separate groups; one with values ranging from 1.5 to 14 hours and the other with values ranging from 19.5 to 24 hours. For Tm_i, the frequency bar at 30 hours represents Tm_i values of 30 hours or larger. As noted above, 30% of the subjects had To_i that were highest at the end of the 24-hour collection period, so it could not be determined if these were actual peak values or not. The majority of the Tm_i were estimated between 1.5 and 8 hours, with a median value of 5.9 hours. Tm_i of 0 hours indicate individuals for whom the nonlinear model did not converge and no parameter estimates were generated.

Figure 3.2. Histogram frequency distribution of urine void time and maximum 1-AP excretion time data following DE exposure in a controlled experiment for observed (*To_i*) and derived (*Tm_i*) results estimated from a cumulative urine excretion model. *Tm_i* larger than 24 hours were set to 30 hours for display purposes.
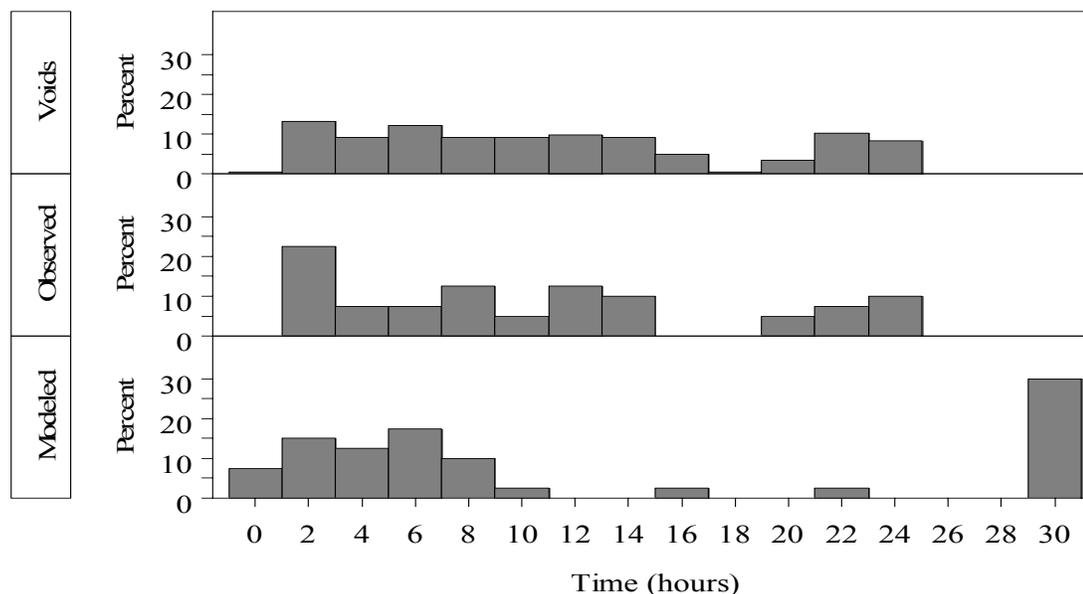
The results of the sensitivity analysis which included all subjects and all samples were similar.

Results from Spot Sampling Simulation

Under the spot-sampling method simulations, the number of subjects providing 4 to 8 post-exposure samples in the simulated data set was similar to that for the DE experiment. Table 3.2 shows the number of subjects and number of samples provided by each.

Table 3.2. Number of Post-exposure Spot Urine Samples in the Simulated Data Sets

Number of Samples	Number of Subjects	Percent
4	7	17.5
5	9	22.5

Number of Samples	Number of Subjects	Percent
6	8	20
7	7	17.5
8	9	22.5

As an illustration of the spot sampling simulation results, a comparative histogram representing the mid-peak and high variability factor selection is provided in Figure 3.3. In this comparative histogram, the void times, $\mathbf{Tmax_i}$, To_i and Tm_i are displayed in rows 1 to 4, respectively. A total of 242 voids from 40 subjects were simulated between 0 and 23.8 hours. The distribution of time points across the 24-hour interval was flat, similar to the experimental data. The $\mathbf{Tmax_i}$ varied between 1.0 and 11.5 hours due to the high simulated inter-subject variability but had a median of 3.62 hours which was similar to the $\mathbf{Tmax_p}$ of 3.7 hours. To_i had the broadest range, from 4.7 to 22.0 hours, with a median of 9.65 hours. This was later than any of the $\mathbf{Tmax_i}$ and did not appear to provide a good estimate of $\mathbf{Tmax_p}$. Tm_i more closely approximated $\mathbf{Tmax_p}$, with a median of 3.70 hours, and a range of (0.6 to 14.0 hours).

Figure 3.3. Histogram frequency distribution of simulated 1-AP urine void time and maximum excretion time data using spot-sampling, mid-level peak, and high parameter variability for Tmax_i, observed *To_i*, and derived *Tm_i* values based on a cumulative urine excretion model.
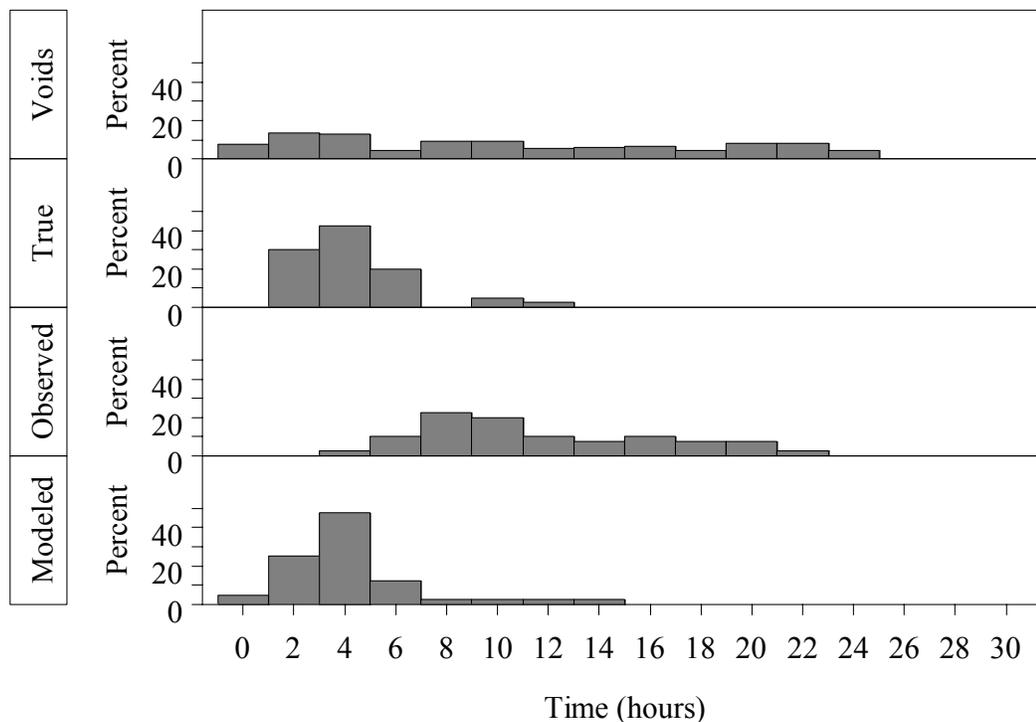
Results from Fixed Sampling Simulation

A total of 320 voids from 40 subjects were simulated between 0 and 24.2 hours in the fixed sampling scheme. This represented approximately one-third more voids than were generated for the spot sampling scheme. Despite the increase in sample collection, summary statistics indicated To_i was not as accurate as Tm_i in estimating **Tmax$_p$**. Review of the summary data for the other factor combinations revealed similar findings (Table 3.3). In all cases, median Tm_i more closely approximated **Tmax$_p$** compared to median To_i. The fixed sampling scheme resulted in improvements in the estimation of **Tmax$_p$** compared to the spot sampling scheme. Increases in parameter variability tended to result in an increase in the range of simulated responses, but had a minimal effect on the median estimate.

Table 3.3. Summary statistics for Tmax$_i$, *To$_i$* and *Tm$_i$* from the experimental and simulated data sets.

Data Source	Factor Levels Variability / Sample Scheme / Tmax$_t$	Tmax$_i$ Median (N=40)	Range (N=40)	*To$_i$* Median (N=40)	Range (N=40)	*Tm$_i$* Median (N=40)	Range (N=40)
Experiment	NA	NA	NA	9.18	1.4-24.2	5.91	0- 30.0
Simulated	Low / Spot / Early	1.50	0.9–2.5	6.90	1.5-13.0	1.51	0-2.4
Simulated	Low / Fixed / Early	1.50	0.9–2.5	3.10	2.4-8.6	1.37	0-2.6
Simulated	High / Spot / Early	1.45	0.5-4.5	6.85	1.4-16.7	1.54	0-4.7
Simulated	High / Fixed / Early	1.45	0.5-4.5	3.10	2.3-12.0	1.38	0.1-4.5
Simulated	Low / Spot / Mid	3.71	2.1-6.2	9.25	4.7-20.8	3.84	0.9-6.6
Simulated	Low / Fixed / Mid	3.71	2.1-6.2	6.00	2.9-12.1	3.69	1.3-6.3
Simulated	High / Spot / Mid	3.62	1.0-11.5	9.65	4.7-22.0	3.70	0.6-14.0
Simulated	High / Fixed / Mid	3.62	1.0-11.5	5.90	2.7-23.7	3.42	0-11.1
Simulated	Low / Spot / Late	8.45	5.0-14.0	16.15	8.6-23.5	8.49	1.5-18.0
Simulated	Low / Fixed / Late	8.45	5.0-14.0	13.65	6.2-24.2	8.42	4.8-13.3
Simulated	High / Spot / Late	8.13	2.6-25.4	16.15	8.6-23.5	8.22	1.4-64.6
Simulated	High / Fixed / Late	8.13	2.6-25.4	23.35	5.5-24.2	8.25	1.9-22.5

Results from Mixture Sampling Simulation

As no single model was identified which would fit the observed 1-AP urine excretion data for the 40 evaluable subjects, no single simulation model was used to mimic the actual data. In order to obtain a simulated excretion data set in which the Tm_i and To_i values mimicked those in the experimental data, a mixture of 3 distributions was employed. One possible explanation for the wide variety of response profiles in the 1-AP excretion data is that more than one type of response profile may exist between subjects. A mixture distribution using spot sampling with high parameter variability and 3 sets of peak times: early, late, and very late, where the very late peak time occurred more than 24 hours after the end of the exposure period was used to generate a simulated data set that would closely approximate the experimental data in terms of To_i and Tm_i. A lag-time model was employed to simulate data for subjects with peak excretion times longer than 24 hours after exposure [Shargel and Yu, 1993].

Figure 3.4. Histogram frequency distribution of simulated 1-AP urine void time and maximum excretion time data using a mixture distribution for void times, Tmax$_i$, *To$_i$* and *Tm$_i$* based on a cumulative urine excretion model.
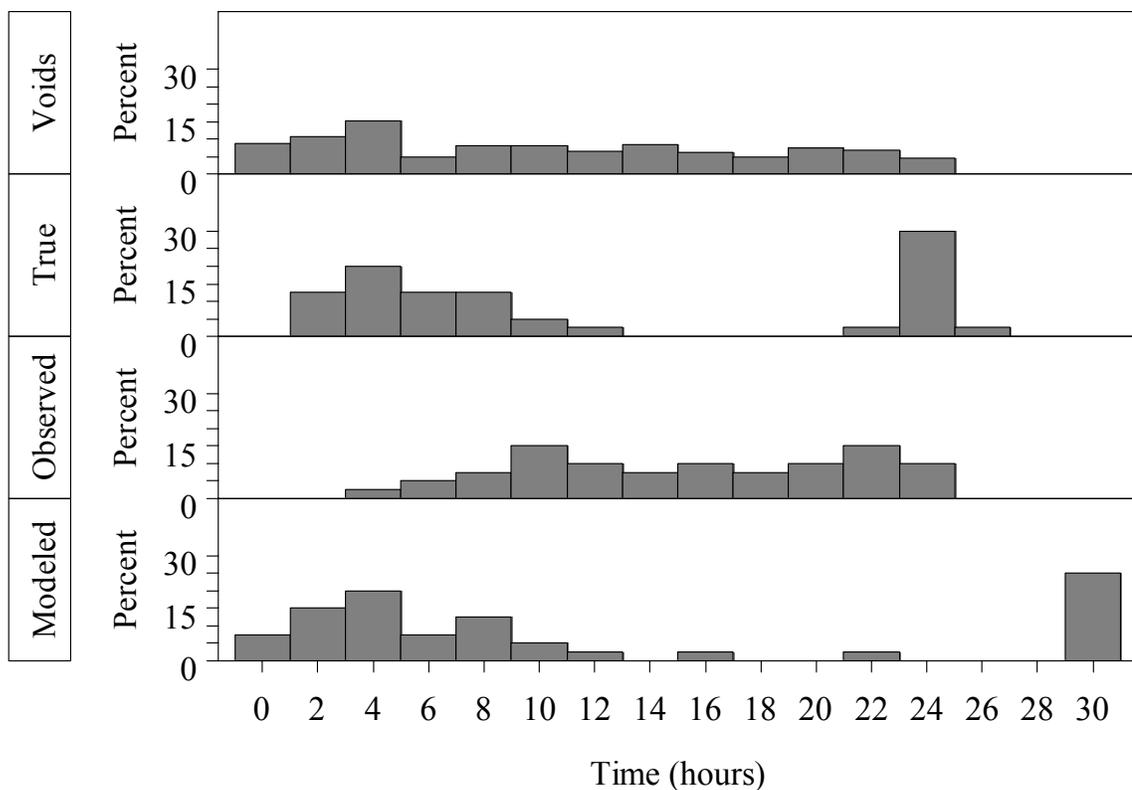
The data distributions for void times, *To$_i$* and *Tm$_i$* (rows 1, 3, and 4 of Figure 3.4) generated by mixing 3 sets of response profiles through simulation closely approximated the experimental data displayed in Figure 3.2. It is therefore hypothesized that the variations in response profiles seen with the experimental 1-AP urine excretion data may be reflective of the presence of a mixture of subject response-types, wherein some individuals may exhibit early peak rates of excretion, while other subjects may exhibit later peak rates.

Discussion

Examination of the DE data revealed a large variety of response profiles. There are many factors that may have contributed to these variations. For example, variations in the rate of formation and elimination of the metabolite, 1-AP from the body may exist between subjects. Studies have shown that the reductive metabolic pathway of 1-NP is mediated by human P450 enzymes [Saito et al., 1984]. Polymorphisms may cause differences in metabolic capacity between individuals that may result in some persons exhibiting slow, intermediate, or rapid biotransformation of xenobiotics such as nitro-PAHs [IPCSINTOXDatabank] . These metabolic differences may be one possible explanation wherein some subjects may be classified as "responders" and some as "non-responders" due to differences in an individual's response to an environmental toxin exposure. Individual responses to an environmental toxin exposure can differ substantially in that some subjects do not exhibit measurable levels of the toxin in their plasma and/or urine despite known exposure (non-responders) and for others (responders) the toxin is present in measurable amounts in their plasma and/or urine. A combination of "responders" and "non-responders" in the subject population could have resulted in a mixture of subject profiles thus accounting for the observed variability in response profiles.

Other possible biological explanations for the presence of different response profiles include differences in respiratory rates as well as variation in routes of absorption. For example, subjects that tend to breathe through their noses might be more likely to experience absorption through the respiratory system, compared to subjects who tend to breathe through their mouths, where absorption through the gastrointestinal system might occur.

These phenomena provide further support for the hypothesis that the observed 1-AP urine excretion profiles may represent a mixture of distributions, perhaps due to subjects exhibiting polymorphisms resulting in different times of maximum 1-AP excretion following a 1-hour controlled exposure period. Future work will explore the identification and characterization of the observed 1-AP urine excretion rates as a mixture distribution using statistical techniques.

Some major challenges existed in the modeling of this DE urine data: variations in response profiles made population-level modeling unobtainable - nearly 30% of the subjects had peak excretion values at the end of the 24-hour period and differed substantially from the remaining subjects; in addition, a number of samples (17%) were declared invalid due to out-of-specification creatinine levels resulting in a potential loss of essential information. It cannot be determined that maximum excretion was obtained for subjects with late-high excretion rates. Further study would be required to evaluate this, for example, a similar study could be conducted, however, with subjects sequestered for up to 48 hours following the controlled exposure period to extend the sampling period to allow for more complete observation of the urine excretion curve for subjects with late peaks and to minimize the chance of secondary exposures. Additional modifications to the data collection process in a future experiment could potentially result in more consistent subject profiles which would increase confidence in the Tmax estimates. For example, the collection of complete urine voids would eliminate the need to standardize the data based on creatinine concentrations and could eliminate some of the variability in response; however, this may not be feasible under some experimental conditions. Nevertheless, in typical pharmacokinetic studies, Tmax is usually determined

experimentally by collecting bio-specimens at pre-defined time intervals. The present analysis demonstrates the utility of spot urine samples collected at irregular time points (according to the natural urinating scheme) in the estimation of Tmax. This presents a large methodological challenge but also a practically useful tool for designing future studies.

Appendix A

This appendix provides details on the data simulation process used in the analysis for this chapter. In all, 13 simulated data sets were created, one for each of the fixed and spot sampling scheme combinations with **Tmax$_p$** and parameter variability factor levels, as well as one for the 3-part mixture distribution. Table A1 provides the factor levels considered in the simulation process.

Table A1. Factor Levels Used to Simulate 1-AP Urine Excretion Profiles

Sampling Schemes	Tmax$_p$ (hours)	Parameter Variability
Fixed	1.5, 3.7, 8.6	low, high
Spot	1.5, 3.7, 8.6	low, high
Three-part Mixed	3.7 & 8.6 & > 24	high

Each simulated data set was generated separately and consisted of 40 subjects as in the evaluable subset of the DE study. In the first step of the simulation process, time points representing urine void times were generated for each subject. Under the fixed sampling paradigm, excretion values were generated to mimic a void every 3 hours from 0 to 18 hours post exposure plus one sample at 24 hours for a total of 8 samples for each subject. Note that the 21-hour time point was not simulated so that an allowance for sleep time was incorporated into the model. When spot sampling was used, subjects were simulated with 4 to 8 urine samples per subject as per the observed data in the DE experiment. The data were generated by first randomly determining the number of time points, n_i, for subject i. Let X_r ($r=1,\ldots, 5$) denote the number of subjects with n_i urine samples where n_i was between 4 and 8, then the number of subjects with n_i urine samples

followed a multinomial distribution where $X_r \sim Multinomial(0.2, 0.2, 0.2, 0.2, 0.2; 40)$. Spot sampling was mimicked by assuming collection times were randomly and uniformly distributed throughout specified periods according to the following distribution:

Table A2. Spot Sampling Time Distribution in the Simulation Model

Sample Number	Time Value Simulation Model
1	Uniform(0, 2)
$>1 - \max(\text{int}(n_i/2))$	Uniform(2, 10)
$>\max(\text{int}(n_i/2)) - n_i - 1$	Uniform(10, 20)
n_i	Uniform(20, 24)

Specifically, each simulated subject provided at least one sample from 0 to 2 hours and another from 20 to 24 hours; the remaining samples were split between the periods from 2 to 10 hours and from 10 to 20 hours post exposure. A fixed seed was used to generate the spot sampling time points to assure the same time point values were generated for each combination of **Tmax$_p$** (early, mid, late) and parameter variability (low, high).

In the second step of the simulation process, cumulative urine concentration values were generated using the cumulative distribution equation below, in a two-stage approach.

$$D_u(t) = \frac{FKeKmD}{Km - K}\left(\frac{1}{K} - \frac{1}{Km}\right) + \frac{FKeKmD}{Km - K}\left[\frac{1}{Km}e^{-Kmt} - \frac{1}{K}e^{-Kt}\right],$$

(1)

In the first stage, parameter values for *K*, *Km*, *Ke*, and *F* were determined for each subject. Parameters followed a multivariate normal distribution specified as:

$$
\begin{bmatrix} \log(K) \\ \log(Km) \\ \log(Ke) \\ \log(F) \end{bmatrix} \sim MVN\left(\begin{bmatrix} \kappa \\ \kappa m \\ \kappa e \\ \phi \end{bmatrix}, \begin{bmatrix} \Sigma_{K,K} & \Sigma_{K,Km} & 0 & 0 \\ \Sigma_{K,Km} & \Sigma_{Km,Km} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right),
$$

where parameters were exponentiated to assure positive estimates. In this parameter distribution, *Ke* and *F* were left as fixed values with no associated variance in order to focus on the parameters associated with Tmax calculation, *K* and *Km*. The expected values for log(*K*), log(*Km*), log(*Ke*) and log(*F*) for each of the planned **Tmax$_p$** values are provided in Table A3 below.

Table A3. Expected Parameter Values in the Simulation Model

Planned Tmax$_p$ (hours)	Parameter			
	κ	κm	κe	ϕ
1.5	-0.5	-0.33	0.6	3.0
3.7	-2.1	-0.7	0.6	3.0
8.6	-2.3	-2.0	0.6	3.0

he   random effects variance/covariance matrices of log(*K*) and log(*Km*) were determined by setting low and high values such that $\Sigma_{K,K}$=0.03 and $\Sigma_{Km,Km}$=0.12 when variability was low, and $\Sigma_{K,K}$=0.15 and $\Sigma_{Km,Km}$=0.6 when variability was high; where $\Sigma_{K,K}$ and $\Sigma_{Km,Km}$ are the random effects variances of log(*K*) and log(*Km*), respectively. The correlation, $\rho_{K,km}$,

was set to 0.04 such that $\Sigma_{K,Km} = \rho_{K,Km} \times \Sigma_{K,K}^{0.5} \times \Sigma_{Km,Km}^{0.5}$; where $\Sigma_{K,Km}$ is the random effects covariance of log(K) and log(Km). The values for low random effects variability and correlation were based on the experimental data results; random effects high variability values were simply set to 5-times the low values. As with the experimental data, D, the amount of DE exposed to the subject, was not included in the modeling as this was standardized as part of the experiment.

In stage two, y_{ij}, the cumulative urine concentration for the ith subject ($i=1,\ldots, m$) at the jth time point, ($j=1,\ldots, n_i$) was calculated as $y_{ij}=Du(t_{ij})+\varepsilon_{ij}$ where the ε_{ij} are the within-subject errors and $\varepsilon_{ij} \overset{iid}{\sim} N(0,21)$. The variance of the ε_{ij} was set using the observed 1-AP excretion data. This resulted in a simulated data set of cumulative urine concentration values with random within- and between-subject variability.

References

Bamford, H.; Baker, J. Nitro-polycyclic aromatic hydrocarbon concentrations and sources in urban and suburban atmospheres of the Mid-Atlantic region. *Atmos. Environ. 2003. 37(15),* 2077-2091.

Barr, D.B.; Wang, R.Y.; Needham,L.L. Biologic Monitoring of Exposure to Environmental Chemicals throughout the Life Stages: Requirements and Issues for consideration for the National Children's Study. *Environ. Health Perspect.* 2005. *113(8),* 1083-1091.

Beland, F. A.; Kadlubar, F. F. Metabolic Activation and DNA Adducts of Aromatic Amines and Nitroaromatic Hydrocarbons. In *Carcinogenesis and Mutagenesis I* ; Cooper, C.S.; Grover, P.L., Eds.; Springer-Verlag: New York, 1990, 267-325.

Bond, J. A., Sun, J. D., Medinsky, M. A., Jones, R. K. , Yeh, H. C. (1986). Deposition, Metabolism, and Excretion of 1-[^{14}C]Nitropyrene and 1-[^{14}C]Nitropyrene Coated on Diesel Exhaust Particles as Influenced by Exposure Concentration. *Toxicology and Applied Pharmacology.* 85, 102-117.

Dutcher, J. S., Sun, J. D., Bechtold, W. E., Unkefer, C. J. (1985). Excretion and Metabolism of 1-Nitropyrene in Rats after Oral or Intraperitoneal Administration. *Toxicological Sciences*, 5 (2), 287-296.

Howard, P. C., Consolo, M. C., Dooley, K. L., Beland, F. A. (1995). Metabolism of 1-Nitropyrene in Mice: Transport Across the Placenta and Mammary Tissues. *Chemico-Biological Interactions,* 95, 309-325.

IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. http://monographs.iarc.fr/ENG/Monographs/vol46/volume46.pdf (accessed July 30, 2008). Volume 46 Diesel and Gasoline Engine Exhausts and Some Nitroarenes.

IPCSINTOXDatabank. http://www.intox.org/databank/documents/chemical/nitrooxy/ehc229.htm. (accessed March 28, 2009).

Laumbach, R. (2009). Quantification of 1-aminopyrene in human urine after a controlled exposure to diesel exhaust**.** *Journal of Environmental Monitoring*, (11) 153-159**.**

Saito, K.; Kamataki, T.; Kato, R. Participation of Cytochrome P-450 in Reductive Metabolism of 1-Nitropyrene by Rat Liver Microsomes. *Cancer Res*. 1984, 44, 3169-3173.

SAS Institute Inc. *SAS OnlineDoc®,* Version 9.1. SAS Institute Inc.: Cary, NC, 2004.

Seidel, A.; Dahmann, D.; Krekeler, H.; Jacob, J. Biomonitoring of polycyclic aromatic compounds in the urine of mining workers occupationally exposed to diesel exhaust. *Int. J. Hyg. Environ. Health.* 2002. *204,* 333-338.

Shargel, L. & Yu, A. B. C. (1993). *Applied Biopharmaceutics and Pharmacokinetics.* Stamford: Appleton & Lang.

Sun, J. D., Wolff, R. K. , Aberman, H. M. , McClellan, R. O. (1983). Inhalation of 1-nitropyrene Associated with Ultrafine Insoluble Particles or as a Pure Aerosol: A Comparison fo Deposition and Biological Fate. *Toxicology Applied Pharmacology,* 69 (2) 185-198.

Tokiwa, H.; Ohnishi, Y. Mutagenicity and carcinogenicity of nitroarenes and their sources in the environment. *Crit Rev Toxicol.* 1986. *17(1),* 23-60.

Chapter 4

Confidence Interval Estimation for Time of Maximum Urine Excretion of Diesel Exhaust Biomarker 1-aminopyrene

Abstract

This chapter provides a comparison of non-parametric and parametric confidence interval techniques for means and percentiles when biomarker concentrations are collected using repeated measure designs. When Tmax is estimated within-subject, confidence intervals for population means and percentiles are derived using standard statistical approaches, including bootstrapping and application of the Central Limit Theorum. When within-subject estimates of Tmax are unreliable or unavailable, first- and second-order delta methods utilizing parameters estimated from nonlinear PK models are used to derive confidence intervals for the mean. These methods are applied to (1) a study of urine 1-aminopyrene, a biomarker for diesel exhaust exposure, (2) a study of theophylline plasma concentrations, and (3) simulated data from both homogeneous and heterogeneous populations. When using within-subject estimates of Tmax from both homogeneous and heterogeneous populations, coverages obtained using the non-parametric method were highest and tended to provide coverages close to the nominal 95% level. When using mean estimates of Tmax, results showed that confidence intervals generated using the first-order delta method provided the highest coverages, albeit below the nominal 95% level at approximately 93% when numerical approximation estimation methods were used.

Introduction

Understanding trends in the time of maximum concentration (Tmax) in pharmacokinetic studies can be critical in characterizing the biological processes surrounding a drug or chemical substance. For example, in pharmaceutical research, characterization of plasma Tmax can be used to plan the dosing schedule of a new drug; and in biomonitoring applications with urine biomarkers, characterization of Tmax can be used to optimize sampling schemes for measuring exposure to environmental pollutants. Here, we seek to characterize the expected time of maximum excretion of 1-aminopyrene, a urine biomarker for diesel exhaust (DE), using data from a recent controlled experiment conducted in human volunteers [Laumbach et al., 2009].

In general, the distribution of Tmax is typically characterized using a measure of central tendency, such as a mean or median, and a measure of dispersion, such as a range or variance. While CI methods for Tmax in the bioequivalence setting have been well documented in the literature [Chow and Liu 1992, Willavize and Morgenthein 2008, Hauschke et al. 1990, Cornell 1991, to list a few], in our review, very little was found about CI methods for Tmax in the single sample case. We suspect that this is due to the context in which pharmacokinetic (PK) studies developed. Specifically, studies of drug availability in the body have typically entailed repeated measures of plasma concentrations in a small number of individuals. These concentrations were frequent, precisely measured and relatively consistent across individuals, so that summaries of the concentrations at each fixed, observed time point were adequate to describe drug availability. Subject-specific Tmax estimates, obtained by selecting the sample time point associated with the highest concentration for each subject, were treated as discrete

and summarized using simple descriptive statistics [Islinger et al. 2006, Cardiello, P. et al. 2000, Smith et al. 2002, Martinez et al. 2007, and Mortimer et al. 2007 as a sample from the literature] or, occasionally, using nonparametric methods based on order statistics or the bootstrap approach [Charles et al., 2007]. We will refer to this as "non-compartmental PK analysis."

We note, however, that the true subject-specific Tmax values could theoretically take any value from a continuous number line and are measured with error. Thus, it is essential to incorporate measures of uncertainty in the estimation within subject and in the estimation of population means when making any conclusions about, for example, average Tmax values. This is particularly true in light of the increased use of biomarkers to measure everything from drug responses to environmental toxin exposure, which may increase the heterogeneity in the distribution of Tmax. Thus, this chapter focuses on creating interval estimates for our estimates of central tendency.

Specifically we focus on creating confidence intervals when either subject-specific or population averages of Tmax are modeled using nonlinear PK modeling, specifically (1) 2-stage nonlinear statistical techniques that initially provide subject-specific estimates that can then be summarized or (2) nonlinear PK random effects models which can directly model population trends including the average Tmax. In the first case, we examine parametric and non-parametric methods for combining information from the subject-specific estimates to create confidence intervals. The latter approach may be particularly appropriate in the case of sparse sampling such as may occur in biomonitoring applications, because there may be insufficient information to allow for estimation of subject-specific Tmax values. In this approach, it may be possible

to directly obtain population-average estimates for PK parameters using nonlinear modeling to calculate the population-average Tmax. However, with this method, no direct estimate of the Tmax standard error is provided to allow for calculation of CIs. Thus, we propose CIs derived using the delta method, which depend on the estimates and standard errors of standard PK parameters such as the absorption and elimination rate constants. Altogether, we compare four CI methods; the nonparametric, bootstrap and standard normal theory approaches to CIs for mean or median Tmax based on individual subject Tmax as well as CIs based on the first- and second-order delta method for the population-average Tmax.

First, we provide an overview of subject-specific and averaged Tmax estimation methods used in this paper, then we present all four approaches to creating the CIs, followed by descriptions of the data utilized in our examples of the CI method application. Specifically, two sample data sets representing different types of pharmacokinetic data will be used to illustrate and compare the CI methods. The first set of data is our primary focus and includes 1-AP concentrations obtained from spot sampled urine voids that represent a sparsely sampled, heterogeneous group of time-concentration profiles. The second consists of theophylline concentrations from serially-collected plasma and represents a richly sampled, homogenous group of time-concentration profiles [SAS Institute, Inc. 2004]. Finally, simulated data sets will be used to evaluate and compare the CI methods.

Materials and Methods

Estimation of Tmax or parameters determining to Tmax

Define **Tmax$_p$** as the population averaged true time to maximum excretion with T_p as its estimate, with some continuous sampling distribution. In addition, each subject i has an individual-specific true maximum time to concentration, **Tmax$_i$** drawn from a population distribution. For that i^{th} subject, T_i ($i=1,...,m$) estimates **Tmax$_i$** with a subject-specific sampling distribution.

To obtain an estimate of the time of maximum excretion, either the values T_i are obtained and then combined to estimate E(**Tmax$_i$**) or a value for T_p is obtained directly and used to estimate E(T_p). In the former case, a non-compartmental or a 2-stage nonlinear technique may be used in the first step; in the latter, population-average nonlinear techniques may be applied. These are as follows:

Non-compartmental approach (NC)

The non-compartmental approach specified the time associated with the highest concentration post-dose or post-exposure as T_i for each subject. We will use the abbreviation "NC" to refer to T_i values based on the non-compartmental method in this chapter. These T_i values may then be used to calculate summary statistics such as mean or median to provide a value for T_p.

Two-stage Nonlinear Approach or Global Two-Stage Method (GTS)

In the first stage, a nonlinear PK model was fit to each subject. Using SAS Proc NLIN, weighted least squares estimates of the parameters were generated through an

iterative process to obtain values that resulted in the smallest error sum-of-squares [Davidian and Giltinan 1995].

An example of a standard nonlinear PK model that could be used in the first stage of the GTS approach is the one-compartment plasma concentration model with first-order rates of absorption and excretion defined as follows: let c_{ij} be the plasma concentration and t_{ij} be the elapsed time since dosing for the ith subject $(i = 1,...,m)$ at the jth post-dose time point $(j = 1,...,n_i)$ and define PK parameters Ka_i, D_i, Ke_i, and Cl_i as the individual subject estimates for absorption rate, dose, excretion rate, and clearance, respectively. Then predicted values would be determined by the following equation [Pinheiro and Bates 2004]:

$$E(c_{ij}) = \frac{Ke_i Ka_i D_i}{Cl_i(Ka_i - Ke_i)} \left[e^{-Ke_i t_{ij}} - e^{-Ka_i t_{ij}} \right] \tag{4.1}$$

Individual GTS Estimates

Following execution of Stage 1 of the GTS approach, subject-specific Ti values were calculated using the maximum likelihood estimate for Ti derived from equation (4.1):

$$g(Ka_i, Ke_i) = \frac{\log(Ka_i / Ke_i)}{(Ka_i - Ke_i)} \tag{4.2}$$

Averaged GTS Estimates

In stage 2 of the GTS approach, average estimates of PK parameters were determined. Several methods have been described for determining these averages, including an EM algorithm approach and a mixed model approach [Davidian and

Giltinan 1995]. In the comparison of CIs for this article, the GTS method with mixed model approach was used to estimate average PK parameters. An alternate method, referred to in the literature as the *Standard Two-Stage* method (STS) [Davidian and Giltinan 2003], consists of calculating sample means and covariance matrices for the PK parameters. In general, the one main distinction between the STS and GTS methods is that the GTS methods have the ability to estimate both within and between subject variances, while the STS method does not. While it is beyond the scope of this article to give a full, detailed description and comparison of these methods, a brief overview of nonlinear mixed modeling statistical theory is provided in Appendix A. The reader is referred to the literature for additional information [Davidian and Giltinan 1993, 1995 and 2003, Lindstrom and Bates 1990, Pinheiro and Bates 1995].

From the mixed model, overall averaged predicted values were obtained for calculation of *Tp* using formulas similar to those for T_i except the subscript 'p' is used to denote the population average estimates, i.e. $g(Ka_p, Ke_p) = \dfrac{\log(Ka_p / Ke_p)}{(Ka_p - Ke_p)}$ for the standard one-compartment plasma concentration model. We will use the abbreviation "GTS" to refer to *Tp* results based on the Global Two-Stage estimation method.

Individual GTSS Estimates

We note that the mixed model used to estimate the overall predicted average PK parameter estimates may also be used to predict individual subject parameter estimates for calculation of T_i. In this approach, the T_i values may differ from the stage 1 T_i values in that the mixed model can separate within and between subject variability. For completeness, we include the T_i provided by the mixed model estimation in the CI

comparisons. We will use the abbreviation "GTSS" to refer to T_i results based on the individual subject predicted values generated from the mixed model analysis.

Population-average Nonlinear Approach (PA)

Nonlinear modeling for the population-average approach utilizes the same PK models as specified above for Stage 1 of the GTS approach. In general, statistical methods utilized in population-average estimation approaches include numerical approximations such as first-order linear approximations and gaussian quadrature approaches to find maximum likelihood estimators for the nonlinear models. Again, it is beyond the scope of this article to provide full detailed descriptions of these various procedures and the reader is referred to several excellent resources for more information [Pinheiro and Bates 1995 and 2004, Wolfinger 1993, Vonesh 1992 and 1993]. In general, these approaches estimate population-average PK parameters along with a between-subject covariance matrix and within-subject estimates of variance. In this paper, SAS Proc NLMixed [SAS Institute Inc., 2004] with the Laplace numerical approximation was used based on the results of a paper by Bates and Pinheiro (1995) demonstrating the advantages of the Laplacian method in terms of accuracy of estimation and based on previous work with 1-AP urine concentration data [Chapters 2 and 3].

Individual and Averaged PA Estimates

Average PK parameters estimated using SAS Proc NLMixed for the population-average model were used to calculate *Tp* as specified above. Predicted values for T_i were also generated and used to compare confidence interval methods for individual subject

data.   We will use the abbreviation "PA" to refer to T_i and Tp results based on the population-average modeling

Individual LAP Estimates

The Laplace numerical approximation method was used to model each subject separately to mimic previous work [Chapters 2 and 3] with the observed 1-AP concentration data.   We will use the abbreviation "LAP" to refer to T_i results based on the Laplace approximation for individual subject modeling.

<u>Reparameterization</u>

Then using nonlinear modeling, it is common practice to reparameterize these PK models using logarithms in order to assure positivity of the PK parameter estimates and make fitting of the model to individual data more stable [Davidian and Giltinan 1995]. Using the standard PK model as an example, and substituting the population-average PK parameters into equation (4.1), define $Ka_p*=\log(Ka_p)$; $Ke_p*=\log(Ke_p)$; etc. for all the PK parameters, then the expected population-average at time j would be defined as:

$$E(c_{pj}) = \frac{e^{Ke_p^*} e^{Ka_p^*} e^{D_p^*}}{e^{Cl_p^*}\left(e^{Ka_p^*} - e^{Ke_p^*}\right)} \left[\exp\left(e^{-Ke_p^* t_{pj}}\right) - \exp\left(e^{-Ka_p^* t_{pj}}\right)\right] \qquad (4.3)$$

and the maximum likelihood estimate for calculation of Tp becomes:

$$g\left(Ka_p^*, Ke_p^*\right) = \frac{\log\left(\exp\{Ka_p^*\}/\exp\{Ke_p^*\}\right)}{\left(\exp\{Ka_p^*\} - \exp\{Ke_p^*\}\right)} = \frac{Ka_p^* - Ke_p^*}{\left(\exp\{Ka_p^*\} - \exp\{Ke_p^*\}\right)}. \qquad (4.4)$$

As stated previously, no value for *Tp* standard error was automatically generated by the GTS or PA methods for the development of CIs. Therefore, standard errors based on first- and second-order delta methods are proposed and derived below.

Confidence Intervals

For quantile estimation, we will treat the observed T_i as though they are the **Tmax$_i$**, that is measured without error. Furthermore, assume the T_i follow a distribution $F(t)$. We examine two methods, (the nonparametric and bootstrap), to create confidence intervals for the $pX100^{th}$ percentile, τ, of the distribution $F(t)$, for example, the population median, and two methods, (the standard normal and delta method), to create a confidence interval for the mean. Note that treating the observed T_i as though they are the **Tmax$_i$** may create some error in the estimates and methods for correcting this have been published. However, as our quantile estimation is focused on the median, we note that these methods for mitigating the effect of the measurement error have been shown to be ineffective in the middle of the distribution [Schechtman and Speigelman 2007]. Therefore, since we primarily focus on the median, we will not apply these corrections.

Nonparametric Method

The nonparametric CI method was applied to T_i values derived from the NC, GTS, GTSS, PA and LAP approaches. A simple approach to determining the confidence interval for the $pX100^{th}$ percentile, τ, is based on nonparametric methods, which make no assumptions about the underlying distribution of the T_i. To find a confidence interval

for the $pX100^{th}$ percentile of a population, begin with the order statistics denoted as $T_{(1)} < T_{(2)} < ... < T_{(m)}$. We wish to find the order statistics $T_{(l)}$ and $T_{(u)}$ that satisfy $P(T_{(l)} < \tau < T_{(u)}) = 1 - \alpha$ where $1 - \alpha$ is the desired probability that the interval captures the $pX100^{th}$ percentile. The interval $[T_{(l)}, T_{(u)}]$ would then constitute the 100(1-α)% confidence interval for τ. Note that for the lth order statistic, $T_{(l)}$, to be less than τ, at least l of the T values must be less than τ. Moreover, for the uth order statistic to be greater than τ, fewer than u of the T values must be less than τ. The probability that at least l and fewer than u of the T values are less than τ is given by the binomial probability for percentile p X100 as $P(T_{(l)} < \tau < T_{(u)}) = \sum_{i=l}^{u-1} \binom{m}{i} (p)^i (1-p)^{m-i}$.

Specifically for the median, where p=0.5, the probability is $\sum_{i=l}^{u-1} \binom{m}{i} (0.5)^m$. To construct a 100(1-α)% confidence interval, choose l and u so that this sum is at least 1-α. To ensure equal-tailed probabilities, l should equal $m - (u - 1)$ when p=0.5; however with smaller sample sizes this may lead to actual confidence intervals that are substantially wider than 100(1-α)%. When working with the median and with a large enough sample size, the normal approximation for the binomial distribution may be used to obtain l and u by solving the following two equations and rounding to the nearest integer:

$$\frac{l - p \times m}{\sqrt{p \times (1-p) \times m}} = -z_{(1-\alpha/2)}, \qquad \frac{u - 1 - p \times m}{\sqrt{p \times (1-p) \times m}} = z_{(1-\alpha/2)}$$

where $z_{(1-\alpha/2)}$ is the $100(1-\alpha/2)^{th}$ percentile of the standard normal distribution [Hogg and Craig 1978].

Bootstrap Method

The Bootstrap CI method was applied to T_i values derived from the NC, GTS, GTSS, PA and LAP approaches. Efron developed methods for constructing approximate confidence intervals using the bootstrap approach (1986, 1987). The bootstrap is a general technique typically used to calculate quantities associated with the sampling distribution of estimators and test statistics [Boos 2006]. It is attractive statistically, as the bootstrap method does not require any distributional assumptions regarding the calculated quantities and yet provides robust confidence intervals under most conditions, i.e. large enough sample size.

In general, given τ is the *px100*th percentile of the individual subject Tmax distribution, a bootstrap "population" can be generated by sampling B times with replacement from the original data values. For each of the B samples, the parameter of interest $\hat{\tau}_b^*$, $b = 1,...,B$ is calculated, where $(*)$ is used to indicate that the calculation arose from the bootstrap "world". Many methods have been developed using the bootstrapped estimates, $\{\hat{\tau}_1^*,...,\hat{\tau}_B^*\}$, to determine confidence intervals. The percentile and bias-corrected methods are described here and both will be used in the assessment of two-sided confidence intervals for Tmax.

The bootstrap percentile method was introduced by Efron in 1979. The percentile method takes the empirical $100(\alpha/2)^{th}$ and $100(1-\alpha/2)^{th}$ percentiles, $0<\alpha<1$, from the bootstrapped values as the lower and upper endpoints of the confidence interval, respectively. In brief, Efron's justification for the percentile interval is based on assuming the existence of an increasing transformation *s* such that

$$P(s(\hat{\tau}) - s(\tau) \le x) = \Phi(x), \tag{4.7}$$

where we assume $\Phi(x)$ is the standard normal distribution function and $\hat{\tau}$ is the estimated value of τ based on the sample. If s is known, an exact lower 2-sided $100(1-\alpha)\%$ confidence bound, (e.g. $\alpha=0.05$), is defined by setting the probability in (4.7) equal to 1-α/2, which results in $x = \Phi^{-1}(1-\alpha/2)$. Then, substituting for x in (4.7) leads to

$$1 - \alpha/2 = P\left(s^{-1}\{s(\hat{\tau}) - \Phi^{-1}(1-\alpha/2)\} \le \tau\right)$$
$$= P\left(s^{-1}\{s(\hat{\tau}) + \Phi^{-1}(\alpha/2)\} \le \tau\right)$$

where $s^{-1}\{s(\hat{\tau}) + \Phi^{-1}(\alpha/2)\}$ is the exact lower 2-sided confidence bound. Similarly an exact upper 2-sided $100(1-\alpha)\%$ bound is defined by $s^{-1}\{s(\hat{\tau}) + \Phi^{-1}(1-\alpha/2)\}$. In the case where s is not known, it can be shown that for $(0 < \alpha < 1)$ where $\hat{\tau}_L^*$ is the lower confidence bound of size $100(1-\alpha)$ and $\hat{\tau}_b^*$ is a random variable based on one resampled bootstrap sample, setting $P_*(\hat{\tau}_b^* \le \hat{\tau}_L^*) = \alpha/2$ leads to

$$P_*\left(s(\hat{\tau}_b^*) - s(\hat{\tau}) \le s(\hat{\tau}_L^*) - s(\hat{\tau})\right) \approx \Phi\left(s(\hat{\tau}_L^*) - s(\hat{\tau})\right)$$

Solving $\alpha/2 = \Phi\left(s(\hat{\tau}_L^*) - s(\hat{\tau})\right)$ for $\hat{\tau}_L^*$ yields $\hat{\tau}_L^* = s^{-1}\{\Phi^{-1}(\alpha/2) + s(\hat{\tau})\}$, the same as the exact bound.

One improvement to the percentile method is the bias-corrected (BC) method which resulted from observing that the estimator under consideration may not be median unbiased, in other words $\hat{\tau} \ne E(\hat{\tau}_{([B+1]/2)}^*)$ where $\hat{\tau}_{([B+1]/2)}^*$ is the median of the bootstrap distribution. Define $H(\cdot)$ as the cdf of $\hat{\tau}_b^*$ where $H(x) = P_*(\hat{\tau}_b^* \le x | \chi)$ and χ is the sample space of resamples. Note that when median bias is present, $H^{-1}(1/2) = \hat{\tau}_{([B+1]/2)}^*$

but $H^{-1}(1/2) \neq \hat{\tau}$. To correct for this median-bias, Efron (1982) allowed for a shift in the

distribution of the parameter estimate, $s(\hat{\tau})$, by a constant, z_0,

$$P\{s(\hat{\tau}) - s(\tau) + z_0 \leq x\} = \Phi(x) \qquad (4.8)$$

Similar to the percentile method, an exact 2-sided lower $100(1-\alpha)\%$ confidence bound for

the bias-corrected bootstrap in the case where s is known is defined as

$s^{-1}\{\Phi^{-1}(\alpha/2) + z_0 + s(\hat{\tau})\}$ which may also be written as $\tau_{ex} = s^{-1}\{z_{\varepsilon/2} + z_0 + s(\hat{\tau})\}$ where

$z_{\varepsilon/2} = \Phi^{-1}(\alpha/2)$ and τ_{ex} is the symbol for the exact lower confidence bound. Using (4.8)

and the cdf $H(\cdot)$ we obtain $H(\hat{\tau}) = P_*(s(\hat{\tau}_b^*) - s(\hat{\tau}) + z_0 \leq z_0) = \Phi(z_0)$. This implies that

$z_0 = \Phi^{-1}(H(\hat{\tau}))$. Note that if $\hat{\tau}$ is median-unbiased, then $P(\hat{\tau} \leq \tau) = \dfrac{1}{2} = \Phi(z_0)$ and

$z_0 = 0$. In the case where s is not known, the lower bound can be determined as follows:

$$\begin{aligned}
1 - \alpha/2 &= \Phi(-z_{\alpha/2}) \\
&= \Phi(s(\hat{\tau}) - s(\tau_{ex}) + z_0) \\
&= P_*\{s(\hat{\tau}^*) - s(\hat{\tau}) \leq s(\hat{\tau}) - s(\tau_{ex})\} \\
&= P_*\{\hat{\tau}^* \leq s^{-1}[2s(\hat{\tau}) - s(\tau_{ex})]\} \\
&= P_*\{\hat{\tau}^* \leq s^{-1}[2s(\hat{\tau}) - (s(\hat{\tau}) + z_0 + z_{\alpha/2})]\} \\
&= P_*\{\hat{\tau}^* \leq s^{-1}[s(\hat{\tau}) - z_0 - z_{\alpha/2}]\}
\end{aligned}$$

which implies that $H^{-1}(1 - \alpha/2) = s^{-1}[s(\hat{\tau}) - z_0 - z_{\alpha/2}]$ which may be written more

generally as:

$$H^{-1}(x) = s^{-1}[s(\hat{\tau}) - z_0 - \Phi^{-1}(x)] \qquad (4.9)$$

By definition of τ_{ex} and (4.9), we obtain the bootstrap bias corrected lower 2-sided

$100(1-\alpha)\%$ confidence bound for τ equal to $H^{-1}(\Phi(z_{\alpha/2} + 2z_0))$ and, using similar

manipulations, the upper 2-sided 100(1-α)% confidence bound equal to $H^{-1}\left(\Phi\left(z_{1-\alpha/2} + 2z_0\right)\right)$ [Boos 2006, Shao&Tu 1995, Hall 1992].

Standard Asymptotic Method

The standard asymptotic CI method was applied to T_i values derived from the NC, GTS, GTSS, PA and LAP approaches. In the standard asymptotic approach, a 2-sided 100(1-α)% confidence interval is determined by

$$\overline{T}_i \pm t_{(1-\alpha/2;m-1)} \times s.e.\left(\overline{T}_i\right) \qquad (4.10)$$

where $\overline{T}_i = \dfrac{1}{m}\sum_{i=1}^{m} T_i$, $s.e.\left(\overline{T}_i\right) = \sum_{i=1}^{m}\left(T_i - \overline{T}_i\right)\Big/(m-1)$, 0<α<1, and $t_{(1-\alpha/2;m-1)}$ is the (1-α/2)

critical value of the Student's t-statistic with m-1 degrees of freedom.

Delta Method

The Delta CI method was applied to T_p values derived from the GTS and PA approaches. The Delta Method gives a normal approximation for the distribution of a statistic which is a function of other statistics that follow a multivariate normal distribution and provides estimates of the expectation and variance of that distribution. Specifically, let $\mathbf{V}_i = \left(V_{1i},...,V_{hi}\right)$ be iid with mean vector $\boldsymbol{\mu} = \left(\mu_1,...,\mu_h\right)$ and covariance matrix Σ, and let $\overline{\mathbf{V}}_h = \left(V_{1h}+,...,+V_{mh}\right)\big/m$. Then, under the central limit theorem,

$$\sqrt{m}\left(\overline{\mathbf{V}}_h - \boldsymbol{\mu}\right) \xrightarrow{D} N\left(0,\Sigma\right).$$

Using the 1-AP urine excretion model as an example, the sampling distribution of T_p is calculated as a function of the K_p^* and Km_p^* where K_p^* and Km_p^* are assumed to

follow a multivariate normal distribution. Suppose $\mathbf{v} = \left(K_p^*, Km_p^*\right)^T$ represent the

population-average parameter estimates based on the fit of the data using the re-

parameterized nonlinear model (5). Given $\left(K_p^*, Km_p^*\right)^T \sim N(\boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{\mu} = (\kappa, \kappa m)^T$,

$$\Sigma = \begin{pmatrix} \sigma_{K_p^*}^2 & \sigma_{K_p^*, Km_p^*} \\ \sigma_{K_p^*, Km_p^*} & \sigma_{Km_p^*}^2 \end{pmatrix}, \quad \text{and} \quad g(\boldsymbol{\mu}) = (\kappa m - \kappa)/(\exp\{\kappa m\} - \exp\{\kappa\}),$$ then using a First

Order Taylor Series expansion, $g(\mathbf{v}) \approx g(\boldsymbol{\mu}) + \dfrac{\partial g(\boldsymbol{\mu})}{\partial \kappa} \times \left(K_p^* - \kappa\right) + \dfrac{\partial g(\boldsymbol{\mu})}{\partial \kappa m} \times \left(Km_p^* - \kappa m\right).$

Given g is a function for which the partial derivatives, $g'(\mathbf{v}) = \dfrac{\partial g(\mathbf{v})}{\partial \mathbf{v}^T}$, are assumed to (1)

exist when $\mathbf{v}$ is evaluated in a neighborhood of $\boldsymbol{\mu}$ and (2) are not all zero, then, as

$m \to \infty$, $\mathrm{E}[g(\mathbf{v})] \to g(\boldsymbol{\mu})$ and $Var[g(\mathbf{v})] \to g'(\boldsymbol{\mu}) \Sigma g'(\boldsymbol{\mu})^T$. Specifically,

$$g(\mathbf{v}) \sim \mathrm{AN}\left[g(\boldsymbol{\mu}), g'(\boldsymbol{\mu}) \Sigma g'(\boldsymbol{\mu})^T\right]$$

with

$$g'(\boldsymbol{\mu})^T = \begin{bmatrix} \dfrac{\partial g(\boldsymbol{\mu})}{\partial \kappa} \\ \dfrac{\partial g(\boldsymbol{\mu})}{\partial \kappa m} \end{bmatrix} = \begin{bmatrix} \dfrac{\exp(\kappa) - \exp(\kappa m) + \exp(\kappa) \times (\kappa m - \kappa)}{[\exp(\kappa m) - \exp(\kappa)]^2} \\ \dfrac{\exp(\kappa m) - \exp(\kappa) - \exp(\kappa m) \times (\kappa m - \kappa)}{[\exp(\kappa m) - \exp(\kappa)]^2} \end{bmatrix}.$$

If we denote $\dfrac{\partial g(\boldsymbol{\mu})}{\partial \kappa}$ as A and $\dfrac{\partial g(\boldsymbol{\mu})}{\partial \kappa m}$ as B, and using matrix multiplication, the variance of

$g(\mathbf{v})$ would be $A^2 \sigma_{K_p^*}^2 + 2AB\sigma_{K_p^*, Km_p^*} + B^2 \sigma_{Km_p^*}^2$ as determined by the First Order Delta

Method.

Here, we propose to use the Delta Method to approximate the sampling distribution of T_p, the population-average estimate of Tmax, however, we note that the distribution of Tmax may be skewed. This classical Delta Method based on a first order Taylor Series expansion of $g(\mathbf{v})$, may not perform well when the function is highly non-linear over the range of values being examined [Cooch and White 2009]. In this case, the Delta Method based on a Second Order Taylor Series expansion may provide a better fit to $g(\mathbf{v})$ by including higher order terms in the expectation and variance formulas. Therefore, $g(\mathbf{v})$ expectation and variance approximations determined by the Second Order Delta Methods were also developed for comparison in this paper. Specifically, using the Second Order Taylor Series expansion with similar assumptions as stated above regarding the function g,

$$g(\mathbf{v}) \approx g(\mathbf{\mu}) + \frac{\partial g(\mathbf{\mu})}{\partial \kappa} \times \left(K_p^* - \kappa\right) + \frac{\partial g(\mathbf{\mu})}{\partial \kappa m} \times \left(Km_p^* - \kappa m\right) +$$

$$\frac{1}{2} \left\{ \frac{\partial^2 g(\mathbf{\mu})}{\partial^2 \kappa} \times \left(K_p^* - \kappa\right)^2 + 2 \times \frac{\partial^2 g(\mathbf{\mu})}{\partial \kappa \partial \kappa m} \left(K_p^* - \kappa\right)\left(Km_p^* - \kappa m\right) + \frac{\partial^2 g(\mathbf{\mu})}{\partial^2 \kappa m} \times \left(Km_p^* - \kappa m\right)^2 \right\}$$

Determination of the expectation and variance for $g(\mathbf{v})$ leads to:

$$E[g(\mathbf{v})] \approx g(\mathbf{\mu}) + \frac{1}{2} \left\{ \frac{\partial^2 g(\mathbf{\mu})}{\partial^2 \kappa} \times \sigma_{K_p^*}^2 + \frac{\partial^2 g(\mathbf{\mu})}{\partial^2 \kappa m} \times \sigma_{Km_p^*}^2 \right\} \text{ and}$$

$$Var[g(\mathbf{v})] \approx Var \left[\begin{array}{c} g(\mathbf{\mu}) + \frac{\partial g(\mathbf{\mu})}{\partial \kappa} \times \left(K_p^* - \kappa\right) + \frac{\partial g(\mathbf{\mu})}{\partial \kappa m} \times \left(Km_p^* - \kappa m\right) + \\ \frac{1}{2} \left\{ \begin{array}{c} \frac{\partial^2 g(\mathbf{\mu})}{\partial^2 \kappa} \times \left(K_p^* - \kappa\right)^2 + \\ 2 \times \frac{\partial^2 g(\mathbf{\mu})}{\partial \kappa \partial \kappa m} \left(K_p^* - \kappa\right)\left(Km_p^* - \kappa m\right) + \frac{\partial^2 g(\mathbf{\mu})}{\partial^2 \kappa m} \times \left(Km_p^* - \kappa m\right)^2 \end{array} \right\} \end{array} \right].$$

This variation (see Appendix C for details of the derivation) is estimated by the following equation:

$$Var[g(\mathbf{v})] \approx \sigma_K^2\left(A^2 + 0.75C^2\right) + \sigma_{Km}^2\left(B^2 + 0.75D^2\right) + \left(\sigma_{K,Km}\right)^2 \times \left(F^2 + CD\right) + F^2\sigma_K^2\sigma_{Km}^2$$
$$- 0.25\left(C^2\sigma_K^4 + D^2\sigma_{Km}^4\right) + 2AB\sigma_{K,Km} + 2CF\sigma_K^2\sigma_{K,Km} + 2FD\sigma_{Km}^2\sigma_{K,Km}$$

with A and B defined as above and

$$C = \frac{\partial^2 g(\mathbf{\mu})}{\partial^2 \kappa} = \frac{\exp(\kappa) \times \{\exp(\kappa m) \times (\kappa m - \kappa - 2) + \exp(\kappa) \times (\kappa m - \kappa + 2)\}}{[\exp(\kappa m) - \exp(\kappa)]^3}$$

$$D = \frac{\partial^2 g(\mathbf{\mu})}{\partial^2 \kappa m} = \frac{\exp(\kappa m) \times \{\exp(\kappa m) \times (\kappa m - \kappa - 2) + \exp(\kappa) \times (\kappa m - \kappa + 2)\}}{[\exp(\kappa m) - \exp(\kappa)]^3}$$

$$F = \frac{\partial^2 g(\mathbf{\mu})}{\partial \kappa \partial \kappa m} = \frac{\exp(\kappa m) \times \{\exp(\kappa m) - 2\kappa m \times \exp(\kappa)\} - \exp(\kappa) \times \{\exp(\kappa) - 2\kappa \times \exp(\kappa m)\}}{[\exp(\kappa m) - \exp(\kappa)]^3}$$

.

The 2-sided $100X(1-\alpha)\%$ confidence interval of Tp was calculated using the standard error, specifically $Tp \pm z_{(1-\alpha/2)} \times \sqrt{\sigma_{Tp}^2/(m-p)}$ where $Tp = g(\mathbf{v})$ and $\sigma_{Tp}^2 = Var[g(\mathbf{v})]$ were generated by both the First Order and Second Order Delta Methods, m is the number of subjects in the trial and p is the number of random effects estimated in the model, and $z_{(1-\alpha/2)}$ is the $100(1-\alpha/2)^{th}$ percentile of the standard normal distribution [Cooch and White 2009, Triantafyllopoulos 2003, Goodman 1962].

EXAMPLE Data Sets

Descriptions of the two sample data sets and the simulated data are provided below.

1-AP Data

This comparison of CI methods and the development of the CI using the Delta Method were motivated by a recent experiment which measured urinary excretion of 1-aminopyrene (1-AP) following a one-hour controlled exposure to diesel exhaust (DE) or clean air (CA) in a two-period randomized cross-over design in healthy volunteers [Lambauch et al, 2009]. 1-AP excreted in the urine is a metabolite of nitro-PAHs, a component of DE particles which has been shown to have genotoxic and carcinogenic properties in experimental animals [IARC Monographs 1989, Tokiwa & Ohnishi 1986, Beland & Kadlubar, 1990]. Thus, this study examined the utility of using 1-AP as a biomarker for DE exposure and found that the cumulative concentration of 1-AP was substantially higher following DE exposure relative to CA. To allow for characterization of the 1-AP excretion-time profile, urine samples were collected over a 24-hour period with timing of samples determined at the convenience of the subjects. However, trends of excretion following the DE exposure proved difficult to characterize due to: (1) the large variation in 1-AP excretion concentrations and in spacing among voids between subjects; and (2) the focus on urine rather than plasma concentrations in that concentrations measured for each urine sample represented a cumulative excreted amount. This, in turn, led to a large amount of uncertainty in the estimates of Tmax for the individual as well as summaries of the Tmax across the population. Despite these difficulties, robust summaries of these trends, perhaps as point and interval estimates for means and percentiles of Tmax in the population would greatly facilitate creation of efficient urine sampling plans that could be used in future observational and experimental studies examining the health effects of DE.

To model the 1-AP data, a modified one-compartment urine excretion model with first-order rates of absorption and excretion was assumed. Cumulative urine concentration values were calculated using the observed data, y^*_{ij}, which were the standardized concentrations from time point $t_{i,j-1}$ to time point t_{ij}. Specifically, if y^*_{ij} is the standardized value for the ith subject ($i=1,\ldots, m$) at the jth time point, ($j=1,\ldots, n_i$) then

$y_{ij} = \sum_{j=2}^{n_i} (t_{ij} - t_{i,j-1}) \times y^*_{ij}$ is the cumulative concentration for subject i from time 0 to time j

where j indexes the post-exposure time points and $y_{ij}=y^*_{i1}$ at $j=1$. Predicted values were determined using estimated parameters derived from fitting a slightly modified one-compartment first-order PK model for total urine excretion given by Equation 4.11 [Shargel and Yu 1993]:

$$E(y_{ij}) = \frac{F_i Ke_i Km_i D_i}{Km_i - K_i}\left(\frac{1}{K_i} - \frac{1}{Km_i}\right) + \frac{F_i Ke_i Km_i D_i}{Km_i - K_i}\left[\frac{1}{Km_i}e^{-Km_i t_{ij}} - \frac{1}{K_i}e^{-K_i t_{ij}}\right] \quad (4.11)$$

Time was defined as the elapsed time in hours since the end of the one-hour controlled DE exposure. The parameters estimated in the nonlinear mixed model included F_i representing the fraction of 1-NP absorbed, Ke_i, the first order renal excretion constant, Km_i, the first order rate constant for metabolism, and K_i, the first order rate constant for elimination. D_i, the amount of DE exposed to the subject, was not a parameter to be estimated because all subjects were exposed to the same level of controlled DE or CA by experimental design. The model modification arose from exchanging Ka_i, the first-order absorption rate constant to Km_i, the rate of change from parent to metabolite in recognition of modeling 1-AP, the metabolite of the parent compound, 1-Nitropyrene. Individual subject-level Tmax values (T_i) and average Tmax values (T_p) were calculated using the following equations:

$$g(Km_i, K_i) = \frac{\log(Km_i / K_i)}{(Km_i - K_i)}$$

and

$$g(Km_p, K_p) = \frac{\log(Km_p / K_p)}{(Km_p - K_p)}$$

Note that the 1-AP urine excretion PK model was re-parameterized as described above for data modeling, however equations using the original parameterization were presented above for simplicity.

Theophylline Data

The theophylline data set [SAS Institute Inc. 2004] contains plasma sample results and is included here to provide an alternate example where more homogenous time-concentration profiles exist between subjects. This experiment consisted of 12 subjects with plasma collection scheduled at one time point prior to administration of a single oral dose of theophylline and ten optimized time points after dosing, up to 25-hours. Subjects were dosed individually on a per-weight basis. The standard one-compartment PK model presented in equation (1) was re-parameterized as described previously and used to fit the theophylline data.

Simulated Data

Three simulated data sets were created to evaluate and compare the CI methods. All three simulated data sets consisted of 1,000 trials with 40 subjects each. Responses were modeled as cumulative urine excretion profiles. The simulations were designed to mimic the diesel exhaust experiment as this was the primary motivation for this research and further understanding of Tmax trends in this setting was desired. In the first

simulation, (Sim1) 4 to 8 spot urine samples per subject were generated using a within-subject variance and between-subject covariance based on the observed DE experiment; however, all subjects were simulated to mimic an excretion rate profile consisting of a single peak. The second simulation (Sim2) was the same as Sim1, except that the correlation in the between-subject covariance matrix was increased from 0.04 to 0.20. The third simulation, (Sim3) was also the same as Sim1, except that instead of simulating spot urine samples, a fixed sampling scheme was used with samples collected every 2 hours over a 24-hour period. Details of the data simulation process are provided in Appendix B.

Data Analysis

Specifically, T_i and Tp values for the theophylline and simulated data sets were generated using each of the 3 approaches; non-compartmental, 2-stage nonlinear modeling, and population-average nonlinear modeling. CIs for each of these approaches were calculated as indicated in Table 4.1 below. Based on previous work in Chapter 3 and due to the sparse sampling and heterogeneity of the data, the non-compartmental and population-average estimation approaches were not viable for estimation of PK parameters in the 1-AP data. However, nonlinear modeling using the Laplace approximation method was able to provide point estimates for individual subject PK parameters which allowed for calculation of T_i in the urine excretion data. Therefore, in the 1-AP data application, CIs were calculated using the nonparametric, bootstrap, and standard normal approaches as applied to the LAP T_i's. In order to compare all CI methods in the urine excretion model, simulated data was used. Summary statistics used to compare the confidence interval methods include the mean and range of the lower and

upper limits. In addition, the simulated data analysis includes the mean CI width, coverage percentage and percent of values above or below the CI. When working with T_i, coverages were determined by calculating the percentage of confidence intervals from the 1000 trials that included the $E(\textbf{Tmax}_i)$ where $E(\textbf{Tmax}_i) =$ $\sum_{i=1}^{N} \textbf{Tmax}_i / N$; $N = \sum_{t=1}^{1000} m_t$ for $t = (1,...,1000)$ simulated trials when the Standard Asymptotic method was used and $E(\textbf{Tmax}_i)=(\textbf{Tmax}_i)_{[(N+1)/2]}$ when the Nonparametric and Bootstrap methods were used. When working with T_p, coverages were calculated based on $E(T_p) = E[g(\textbf{v})]$ previously defined in the delta method derivations.

Table 4.1: Confidence Interval Assessment

Tmax Method(s)	Data Set(s)	Confidence Interval Methods
NC / GTS / GTSS (T_i)	Theophylline, Sim1, Sim2, Sim3	Nonparametric Bootstrap Standard Asymptotic
LAP (T_i)	1-AP, Theophylline, Sim1, Sim2, Sim3	Nonparametric Bootstrap Standard Asymptotic
PA (T_i)	Theophylline, Sim1, Sim2, Sim3	Nonparametric Bootstrap Standard Asymptotic
GTS / PA (T_p)	Theophylline, Sim1, Sim2, Sim3	Delta Method

Results

Observed Data

Fifty-five healthy volunteers participated in the DE exposure study. This group consisted of 33 males and 22 females with an average age of 24.8 years (std dev=6.59). Fourteen subjects had incomplete or invalid data following the DE exposure arm and one

subject discontinued prior to completing the study. These 15 subjects were excluded

from the analysis. The 40 analyzed subjects consisted of 25 males and 15 females with

an average age of 24.0 years (std dev=5.59), similar to the all-randomized group.

Previous analysis [Chapter 3] reviewed the poolability of the DE observations obtained

from the two cross-over periods and found no issues of concern.

Table 4.2: 1-AP Tmax Confidence Interval Estimate (hours) for 25 Subjects Exposed to Diesel Exhaust with Mean Ti = 6.04 hours and Median Ti = 5.37 hours Based on the LAP Modeling Approach.
.

CI Method	CI (Range)
Non-Parametric	3.91– 6.22 (2.31)
Bootstrap: Percentile	3.91 – 6.22 (2.31)
Bootstrap: Bias-corrected	3.33– 5.60 (2.27)
Standard Asymptotic	4.18 – 7.90 (3.72)

Of the 40 subjects available for analysis, 15 subjects had data which did not

converge using the LAP method. These subjects were therefore excluded from the

calculation and evaluation of CIs, leaving a total of 25 subjects. The 1-AP T_i distribution

in the DE-exposed subjects was not symmetric and resulted in a median T_i of 5.37 hours

and mean T_i of 6.04 hours. The non-parametric and bootstrap methods yielded similar

confidence intervals as shown in Table 4.2. The widest confidence intervals were

produced by the standard asymptotic method at 4.18-7.90 hours, with a range that was

approximately 1 hour longer than the ranges associated with the non-distributional methods.

Theophylline Data

Table 4.3: Tmax Confidence Interval Estimate (hours) for 12 Subjects Exposed to Theophylline

.

		Modeling Approaches				
		NC	GTSS	GTS	LAP	PA
Mean T_i		1.79	*2*.31	2.34	2.10	2.08
Median T_i		1.14	2.43	2.41	2.07	2.07
T_p		NA	NA	2.15	NA	1.93
Non-Parametric	**CI (Range)**	1.00–3.48 (2.48)	1.50-2.98 (1.48)	1.51-3.37 (1.86)	1.44-2.84 (1.40)	1.48–2.73 (1.25)
Bootstrap: Percentile	**CI (Range)**	1.01–2.75 (1.74)	1.66-2.82 (1.16)	1.73–3.11 (1.38)	1.52-2.62 (1.10)	1.55–2.53 (0.98)
Bootstrap: Bias-corr.	**CI (Range)**	1.00– 2.32 (1.32)	1.66-2.81 (1.15)	1.73–3.04 (1.31)	1.44-2.39 (0.95)	1.55–2.42 (0.94)
Standard Asymptotic	**CI (Range)**	1.08–2.50 (1.42)	1.74-2.89 (1.15)	1.70–2.98 (1.28)	1.53-2.66 (1.13)	1.57–2.59 (1.02)
1st Order Delta	**CI (Range)**	NA	NA	1.57–2.72 (1.15)	NA	1.44–2.42 (0.98)
2nd Order Delta	**CI (Range)**	NA	NA	1.51-2.79 (1.28)	NA	1.37-2.49 (1.12)

Due to the homogeneity of the theophylline data, it was possible to apply different modeling approaches as well as the various CI methods to the data. Results are provided in Table 4.3. Calculation of mean and median T_i as well as Tp varied among the modeling approaches. The lowest mean and median T_i occurred using the NC approach and the highest using the GTS and GTSS methods where median T_i 's were double the value the NC median T_i. The Tp value for the GTS approach was also longer than the Tp

value calculated using the PA method estimates. Among the methods, GTS and GTSS results were similar to each other and the LAP results were similar to the PA methods. Among the non-distributional CI methods, the non-parametric method provided the widest CI while the bias-corrected bootstrap had the narrowest range. Between the Delta methods, the CI range for the first order Delta method was narrower than the CI range for the second order for both the GTS and PA modeling approaches. It is difficult to ascertain which modeling approach and which CI method are the most reliable as we don't know the true value of the time of maximum concentration for the theophylline data, therefore, we look to data simulations to evaluate this.

Simulated Cumulative 1-AP Urine Excretion Data

One-thousand simulated "trials" with 40 subjects each were used not only to compare the Tmax confidence interval approaches but also the modeling methods for each of the three sets of simulated data. As we know the true parameter values, we can evaluate their accuracy and precision; we begin with an evaluation of the modeling approaches.

Figure 4.1 below is a series of histograms displaying the distribution of the Sim1 $\mathbf{Tmax_i}$ values in row 1, and the NC T_i , GTS T_i , GTSS T_i , LAP T_i and PA T_i in rows 2 through 6. Confirming earlier work [Chapter 3], the NC approach for urine excretion data was the least accurate for estimating time of maximum urine excretion and would likely lead to a biased estimation of E($\mathbf{Tmax_i}$), higher than the true value. The numerical approximation (LAP and PA) T_i values most closely mimicked those of the $\mathbf{Tmax_{i.,}}$ Although the GTS and GTSS T_i values were approximately centered near the $\mathbf{Tmax_p}$ of 3.74 hours, they were more variable than the LAP and PA T_i. GTSS and PA T_i values

obtained during processing of the "trial" as a whole were less variable than values obtained during by-subject processing of the data, e.g. LAP and GTS T_i values. Convergence difficulties were encountered with the GTS and LAP methods, likely due to the sparseness of the data, but which resulted in approximately 5% of the subjects' data being excluded from the GTS analysis and approximately 2% of the subject's data being excluded from the LAP analysis. Patterns in T_i distribution were similar in the Sim2 and Sim3 data sets; the corresponding histograms are provided in Appendix D.
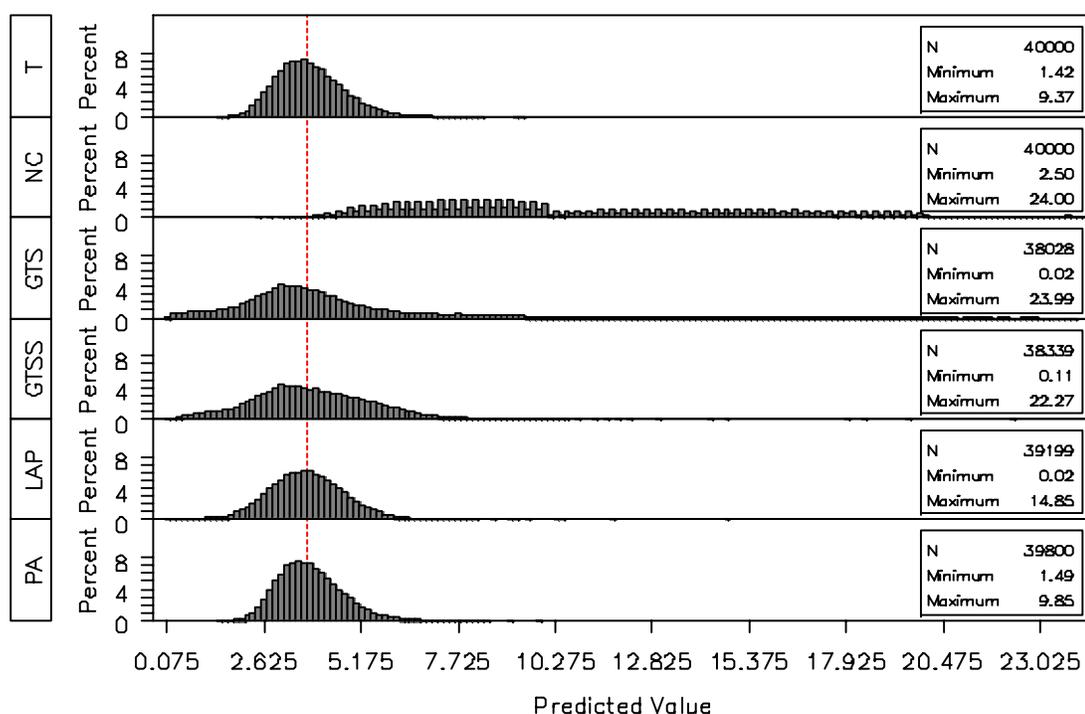
Figure 4.1: Distribution of Sim1 T_i Values [rows 2 to 6] as compared to the Tmax$_i$ values [row 1], with a reference line indicating Tmax$_p$ of 3.74 hours.

Table 4.4 presents summary statistics comparing the T_p values, and K_p* and Km_p* estimates generated using the GTS and PA nonlinear modeling approaches with the "true" values used to generate the simulated data for the Sim1, Sim2, and Sim3 data sets. While both modeling methods provided similarly accurate estimates in terms of the mean

K_p*, the mean T_p and Km_p* provided by the GTS method appeared to underestimate the "true" values. In addition, the GTS method resulted in more variability in T_p, K_p*, and Km_p* and had a greater bias, showing a tendency to underestimate **Tmax$_p$.** Overall, the PA method was more accurate and more precise than the GTS in generating values for T_p, K_p*, and Km_p*. Results for each of the three simulations were similar.

Table 4.4: **Comparison of Calculated and Estimated Parameters with "True" Values Based on Simulated Data for n=1000 "trials" for Sim1, Sim2, and Sim3 Data**

	2-Stage (GTS) Nonlinear Modeling			Population-Average Nonlinear Modeling		
	Tmax	K_p*	Km_p*	Tmax	K_p*	Km_p*
True	3.74	-2.1	-0.7	3.74	-2.1	-0.7
Simulation 1						
Mean	3.582	-2.108	-0.618	3.744	-2.100	-0.700
Min	2.674	-4.027	-0.972	3.238	-2.244	-0.941
Max	6.452	-1.980	-0.181	4.286	-1.961	-0.480
Var	0.143	0.024	0.015	0.027	0.002	0.006
Bias	-0.158	-1.408	1.482	0.004	-1.400	1.400
MSE	0.168	2.007	2.213	0.027	1.962	1.967
Simulation 2						
Mean	3.583	-2.108	-0.618	3.744	-2.100	-0.700
Min	2.684	-3.419	-1.052	3.199	-2.246	-0.944
Max	6.577	-1.982	-0.189	4.313	-1.961	-0.466
Var	0.148	0.022	0.015	0.028	0.002	0.006
Bias	-0.157	-1.408	1.482	0.004	-1.400	1.400
MSE	0.173	2.004	2.211	0.028	1.962	1.967
Simulation 3						
Mean	3.561	-2.091	-0.615	3.741	-2.100	-0.699
Min	2.460	-3.727	-0.989	3.277	-2.344	-1.023
Max	6.624	-1.879	-0.092	4.225	-1.842	-0.422
Var	0.239	0.051	0.018	0.025	0.004	0.008
Bias	-0.179	-1.391	1.485	0.001	-1.400	1.401
MSE	0.271	1.984	2.221	0.025	1.965	1.970

A histogram comparing the distribution of the GTS T_p and PA T_p values for the 1000 simulated trials in the Sim1 data is displayed in Figure 4.2. The GTS results in row

1 show a tendency to underestimate **Tmax$_p$** and portray a right skew, while the PA results were centered on **Tmax$_p$** and show a bell-shaped, symmetric distribution.
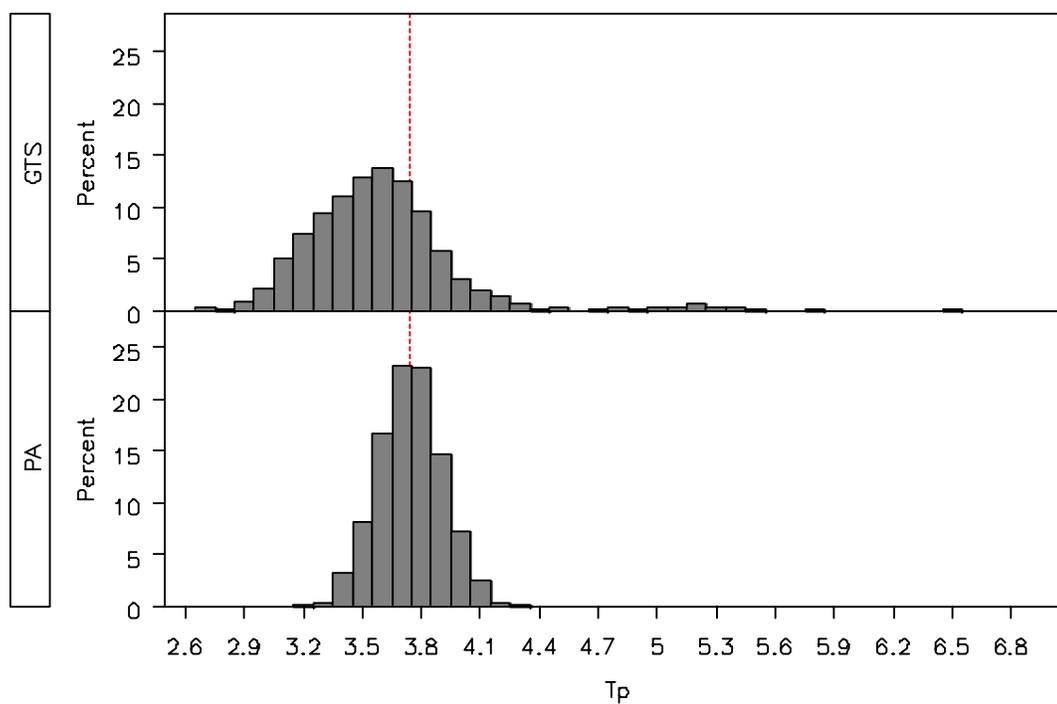
Figure 4.2: Sim 1 T_p calculated with 2-stage (GTS) and Population-average (PA) nonlinear modeling estimates, reference line indicates Tmax$_p$ value of 3.74 hours.

In addition to providing values for **Tmax$_p$**, κ, and κm in the simulations, we set values for the random effects variance covariance matrices, $\sum_{K,K}$, $\sum_{Km,Km}$, and $\sum_{K,Km}$. The histograms in Figures 4.3 to 4.5 below display a comparison of the estimated $\sum_{K,K}$, $\sum_{Km,Km}$, and $\sum_{K,Km}$ provided by the GTS and PA methods using the Sim1 data. Simulated trial results with out-of-trend values were not displayed in the histograms. A comparison of the estimated values for variance and covariance with the "true" values is provided in Table 4.5. Estimates of $\sum_{Km,Km}$, and $\sum_{K,Km}$ obtained by the GTS nonlinear

modeling method were inaccurate and tended to over-estimate the "true" values for $\sum_{Km,KM}$ and $\sum_{K,Km}$ while the estimates provided by the PA method tended to be centered around the true values showing less bias and variability (Table 4.5). The Sim2 and Sim3 data sets displayed similar trends in variance over-estimation, see Appendix D for histogram displays.
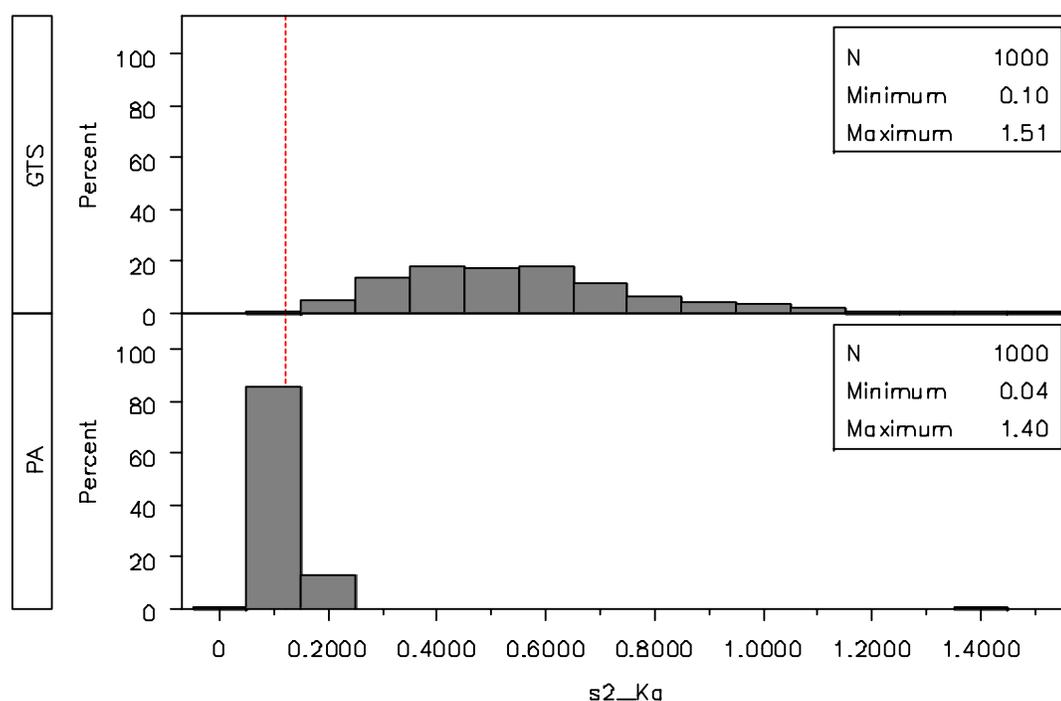
Figure 4.3: Sim 1 $\sum_{Km,Km}$ estimated by 2-stage (GTS) and Population-average (PA) nonlinear modeling, reference line indicates "true" value of 0.12.
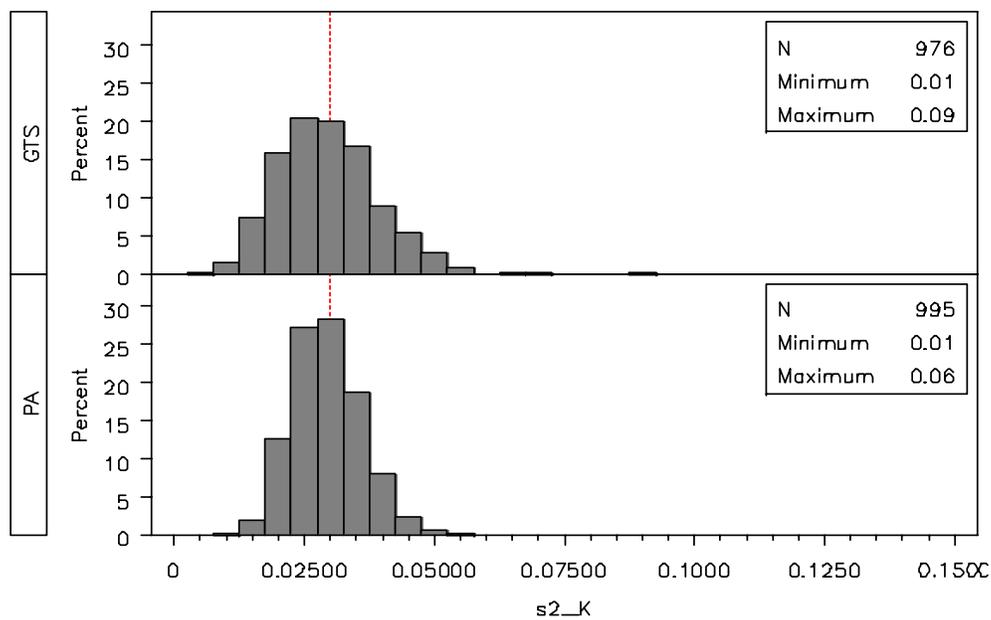
Figure 4.4: Sim 1 $\sum_{K,K}$ estimated by 2-stage (GTS) and Population-average (PA) nonlinear modeling, reference line indicates "true" value of 0.03.
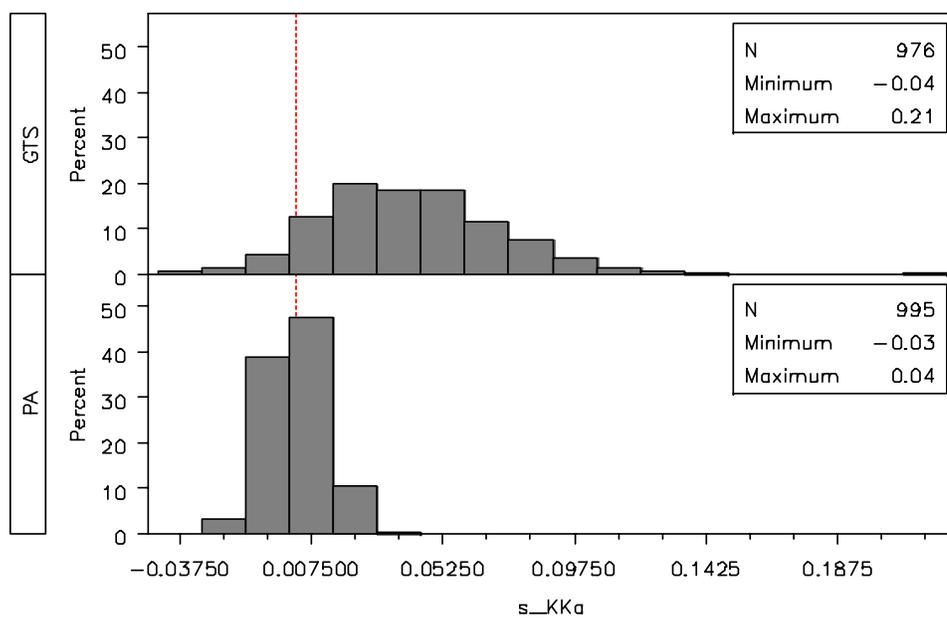
Figure 4.5: Sim 1 $\sum_{K,Km}$ estimated by 2-stage (GTS) and Population-average (PA) nonlinear modeling, reference line indicates "true" value of 0.0024.

Table 4.5: Comparison of Estimated Variance and Covariance Parameters with "True" Values Based on Simulated Data for n=1000 "trials" for Sim1, Sim2, and Sim3 Data

	2-Stage Nonlinear Modeling			Population-Average Nonlinear Modeling		
	$\sum_{K,Km}$	$\sum_{K,K}$	$\sum_{Km,Km}$	$\sum_{K,Km}$	$\sum_{K,K}$	$\sum_{Km,Km}$
Simulation 1						
True Value	0.0024	0.03	0.12	0.0024	0.03	0.12
Mean	0.0846	0.7571	0.5494	0.0063	0.0356	0.1227
Min	-0.0380	0.0073	0.0985	-0.0279	0.0110	0.0408
Max	3.6793	60.8939	1.5133	0.8000	1.3000	1.4000
Var	0.0829	23.0565	0.0500	0.0033	0.0081	0.0091
Bias	0.0822	0.7271	0.4294	0.0039	0.0056	0.0027
MSE	0.0897	23.5852	0.2344	0.0033	0.0081	0.0091
Simulation 2						
True Value	0.012	0.03	0.12	0.012	0.03	0.12
Mean	0.0974	0.7264	0.5485	0.0195	0.0419	0.1293
Min	-0.0294	0.0071	0.0864	-0.0159	0.0110	0.0442
Max	3.0813	36.7754	2.5213	0.8000	1.3000	1.4000
Var	0.0886	20.4244	0.0539	0.0063	0.0160	0.0173
Bias	0.0854	0.6964	0.4285	0.0075	0.0119	0.0093
MSE	0.0959	20.9094	0.2375	0.0063	0.0162	0.0173
Simulation 3						
True Value	0.0024	0.03	0.12	0.0024	0.03	0.12
Mean	0.1333	1.1996	0.5238	0.0064	0.0359	0.1239
Min	-0.0283	0.0019	0.1178	-0.0417	0.0121	0.0443
Max	2.3357	41.9359	1.6933	0.8000	1.3000	1.4000
Var	0.0774	34.0640	0.0398	0.0033	0.0081	0.0091
Bias	0.1309	1.1696	0.4038	0.0040	0.0059	0.0039
MSE	0.0946	35.4320	0.2029	0.0033	0.0081	0.0091

Despite these challenges, evaluation of confidence intervals for T_p was desired, both to help in understanding the data collected and for designing future studies in terms of timing of sample collection. Tables 4.6 through 4.8 present summaries of the confidence interval method comparisons for each of the Tmax estimation approaches using the three simulated data sets, respectively. Individual subject data that did not converge in the nonlinear modeling process was not included in the calculation of CIs. As stated above, this accounted for approximately 5% of the GTS subject-specific evaluations and approximately 2% of the LAP subject-specific evaluations. Overall, confidence interval method characteristics such as coverage trends and widths were similar across the simulations indicating that the differences in sampling schemes (convenience vs. fixed) and the correlation ($\rho_{K,Km}$) did not impact the relative performance of the CI methods.

We begin with a review of the CI methods based on the *Ti* values. In general, we note that the CIs based on the NC approach were too high and provided zero coverage of E(**Tmax**$_i$). The standard asymptotic CIs were consistently the least accurate and tended to provide the lowest coverage regardless of the nonlinear estimation method or simulation. The non-parametric method, on the other hand, consistently performed the best, providing the highest coverage in each set of simulation and nonlinear modeling groupings for CIs calculated from the *Ti* values with coverage values ranging from 91.4% to 95.3%. While the various CI methods resulted in similar coverages for the GTSS, GTS, and LAP methods, the widths of the confidence intervals tended to be narrowest for *Ti* values based on the LAP estimation approach. The PA *Ti* values were associated with lower coverages for each of the confidence interval methods.

The first and second order delta methods were the only CI methods available for *Tp* when directly calculated using estimates provided by the GTS and PA methods. In both the GTS and PA modeling approaches in each of the 3 simulations, the CIs generated using the first order delta method provided higher coverage, although all coverages were lower than the nominal 95% level. Coverages provided by the first order delta method CIs ranged from 92.7% to 93.4% for the PA method and from 83.7% to 89.0% for the GTS method. As with the *Ti* confidence intervals, the mean CI widths based on *Tp* values calculated from GTS method-derived parameters were approximately double in size and more variable compared to widths based on the PA method-derived parameters.

Table 4.6: Comparison of Confidence Interval Methods Based on Sim1 Data

Tmax Estimation Method: Parameter	CI Method	Lower Limit Mean Min - Max	Upper Limit Mean Min - Max	Mean Width (std dev)	Coverage (%) [%below, %above]
NC: *Ti*	Non-Parametric	7.99 6.0-11.2	12.33 8.8-17.2	4.34 (0.28)	0 [0, 100]
	Bootstrap: Percentile	8.10 6.15-11.20	12.07 8.75-16.80	3.97 (1.25)	0 [0, 100]
	Bootstrap: Bias-corrected	7.98 6.00-11.10	11.76 8.70-16.40	3.78 (1.28)	0 [0, 100]
	Standard Asymptotic	9.32 7.36-11.36	12.21 9.74-14.44	2.89 (0.28)	0 [0, 100]
GTSS[a]: *Ti*	Non-Parametric	3.15 2.24-4.26	4.42 3.45-5.78	1.26 (0.34)	94.8 [2.0, 3.2]
	Bootstrap: Percentile	3.18 2.32-4.26	4.38 3.45-5.67	1.20 (0.32)	94.2 [2.5, 3.3]
	Bootstrap: Bias-corrected	3.12 1.96-4.26	4.30 3.39-5.65	1.18 (0.32)	93.4 [4.4, 2.2]
	Standard	3.34	4.31	0.97	92.0

	Asymptotic	2.53-4.25	3.45-5.76	(0.18)	[3.9, 4.1]
GTS^b: T_i	**Non-Parametric**	3.12 2.16 - 4.18	4.49 3.49 – 6.77	1.37 (0.41)	94.9 [2.4, 2.7]
	Bootstrap: Percentile	3.15 2.31 – 4.18	4.45 3.47-6.77	1.30 (0.39)	94.6 [2.8, 2.6]
	Bootstrap: Bias-corrected	3.09 1.92 – 4.07	4.34 3.38 – 6.51	1.25 (0.37)	92.6 [5.3, 2.1]
	Standard Asymptotic	3.48 2.59-4.67	5.46 3.65 – 8.04	1.98 (.50)	84.5 [0.1, 15.4]
PA: T_i	**Non-Parametric**	3.42 2.92-3.99	4.09 3.54-4.84	0.68 (0.17)	91.2 [3.8, 5.0]
	Bootstrap: Percentile	3.44 2.92-4.05	4.06 3.50-4.83	0.62 (0.16)	88.9 [4.9, 6.1]
	Bootstrap: Bias-corrected	3.43 2.92-4.03	4.04 3.50-4.83	0.61 (0.16)	88.6 [5.9, 5.4]
	Standard Asymptotic	3.56 3.08-4.07	4.08 3.55-4.70	0.52 (0.08)	86.9 [7.0, 6.0]
LAP^c: T_i	**Non-Parametric**	3.30 2.71-3.92	4.11 3.41-4.81	0.81 (0.20)	95.5 [2.3, 2.2]
	Bootstrap: Percentile	3.32 2.71-3.92	4.08 3.40-4.72	0.76 (0.18)	94.0 [3.3, 2.7]
	Bootstrap: Bias-corrected	3.28 2.69-3.92	4.04 3.39-4.65	0.76 (0.18)	92.0 [5.7, 2.3]
	Standard Asymptotic	3.36 2.74-4.05	4.04 3.52-4.69	0.68 (0.09)	89.3 [9.8, 0.9]
GTS: T_p	**First Order Delta Method**	2.99 1.33-4.07	4.18 3.20-11.57	1.19 (0.89)	87.7 [11.6, 0.7]
	Second Order Delta Method	2.92 -0.83-4.04	4.24 3.24-13.73	1.32 (1.20)	60.6 [39.2, 0.2]
PA: T_p	**First Order Delta Method**	3.44 2.92-3.94	4.04 3.52-4.63	0.60 (0.06)	93.3 [3.5, 3.2]
	Second Order Delta Method	3.41 2.89-3.91	4.07 3.54-4.66	0.66 (0.08)	89.2 [10.5, 0.3]

[a] Not all Sim1 subjects converged in the GTSS method, although approximately 98% of the trials had PK parameter estimates from 35 or more subjects.

[b] Not all Sim1 subjects converged in the GTS method, although approximately 98% of the trials had PK parameter estimates from 35 or more subjects.

[c] Not all Sim1 subjects converged in the LAP method, although all of the trials had PK parameter estimates from 35 or more subjects.

Table 4.7: Comparison of Confidence Interval Methods Based on Sim2 Data

Tmax Estimation Method	CI Method	Lower Limit Mean Min - Max	Upper Limit Mean Min - Max	Mean Width (std dev)	Coverage (%) [%below, %above]
NC: T_i	Non-Parametric	8.00 6.00-11.20	12.36 8.80-17.20	4.36 (1.31)	0 [0, 100]
	Bootstrap: Percentile	8.11 6.15-11.25	12.10 8.80-16.60	4.00 (1.26)	0 [0, 100]
	Bootstrap: Bias-corrected	7.98 5.70-11.10	11.79 8.40-16.40	3.81 (1.27)	0 [0, 100]
	Standard Asymptotic	9.34 7.36-11.36	12.25 9.74-14.44	2.91 (0.27)	0 [0, 100]
GTSS[a]: T_i	Non-Parametric	3.14 2.25-4.23	4.44 3.42-6.07	1.30 (0.34)	94.1 [2.2, 3.7]
	Bootstrap: Percentile	3.16 2.25-4.23	4.40 3.42-5.71	1.24 (0.33)	93.2 [2.6, 4.2]
	Bootstrap: Bias-corrected	3.11 1.93-4.20	4.32 3.10-5.71	1.21 (0.32)	93.5 [4.0, 2.5]
	Standard Asymptotic	3.35 2.54-4.30	4.33 3.49-5.74	0.98 (0.17)	91.5 [4.2, 4.3]
GTS[b]: T_i	Non-Parametric	3.11 2.17-4.07	4.51 3.52-6.93	1.39 (0.41)	94.7 [2.2, 3.1]
	Bootstrap: Percentile	3.15 2.25 – 4.15	4.46 3.48-6.93	1.32 (0.39)	94.0 [2.6, 3.4]
	Bootstrap: Bias-corrected	3.08 1.92-3.99	4.36 3.37-6.93	1.28 (0.38)	92.7 [5.0, 2.3]

	Standard Asymptotic	3.48 2.61-4.69	5.48 3.69-8.25	2.00 (0.51)	84.9 [0.1, 15.0]
PA: T_i	**Non-Parametric**	3.41 2.90-4.10	4.11 3.53-4.95	0.70 (0.18)	95.6 [2.6, 1.8]
	Bootstrap: Percentile	3.43 2.90-4.10	4.08 3.47-4.85	0.65 (0.17)	94.7 [3.4, 1.9]
	Bootstrap: Bias-corrected	3.41 2.87-4.10	4.05 3.48-4.86	0.64 (0.16)	92.9 [5.8, 1.3]
	Standard Asymptotic	3.56 3.04-4.09	4.11 3.52-4.77	0.55 (0.08)	88.2 [6.9, 5.0]
LAP^c: T_i	**Non-Parametric**	3.29 2.71-3.97	4.13 3.46-4.80	0.84 (0.21)	95.3 [3.2, 1.5]
	Bootstrap: Percentile	3.31 2.77-3.97	4.10 3.44-4.75	0.79 (0.20)	94.1 [4.1, 1.8]
	Bootstrap: Bias-corrected	3.28 2.69-3.96	4.06 3.40-4.73	0.78 (0.20)	91.3 [7.2, 1.5]
	Standard Asymptotic	3.36 2.73-4.04	4.06 3.53-5.43	0.70 (0.12)	89.3 [10.0, 0.7]
GTS: Tp	**First Order Delta Method**	2.98 1.41-3.94	4.18 3.21-11.17	1.20 (0.89)	87.2 [12.1, 0.7]
	Second Order Delta Method	2.92 -0.83-3.91	4.25 3.25-12.46	1.33 (1.21)	61.2 [38.7, 0.1]
PA: Tp	**First Order Delta Method**	3.43 2.90-3.95	4.06 3.48-4.68	0.62 (0.06)	92.7 [3.9, 3.3]
	Second Order Delta Method	3.40 2.87-3.92	4.08 3.50-4.71	0.68 (0.07)	89.2 [10.5, 0.3]

[a] Not all Sim2 subjects converged in the GTSS method, although approximately 99% of the trials had PK parameter estimates from 35 or more subjects.

[b] Not all Sim2 subjects converged in the GTS method, although approximately 98% of the trials had PK parameter estimates from 35 or more subjects.

[c] Not all Sim2 subjects converged in the LAP method, although all of the trials had PK

parameter estimates from 35 or more subjects.

Table 4.8: Comparison of Confidence Interval Methods Based on Sim3 Data

Tmax Estimation Method	CI Method	Lower Limit Mean Min - Max	Upper Limit Mean Min - Max	Mean Width (std dev)	Coverage (%) [%below, %above]
NC: T_i	Non-Parametric	4.90 3.80-5.90	6.30 5.50-7.90	1.39 (0.78)	0 [0, 100]
	Bootstrap: Percentile	4.99 3.80-5.90	6.22 5.40-7.95	1.23 (0.70)	0 [0, 100]
	Bootstrap: Bias-corrected	4.72 3.75-5.90	6.04 4.00-7.80	1.33 (0.65)	0 [0, 100]
	Standard Asymptotic	5.27 4.33-6.35	6.50 5.20-7.67	1.22 (0.15)	0 [0, 100]
GTSS[a]: T_i	Non-Parametric	3.10 2.12-4.50	4.50 3.20-6.83	1.42 (0.40)	92.8 [2.0, 5.1]
	Bootstrap: Percentile	3.13 2.12-4.50	4.48 3.20-6.83	1.36 (0.40)	91.7 [2.3, 6.0]
	Bootstrap: Bias-corrected	3.05 2.00-4.50	4.41 3.07-6.45	1.36 (0.38)	92.1 [3.8, 4.0]
	Standard Asymptotic	3.30 2.54-4.54	4.31 3.45-7.62	1.01 (0.25)	88.9 [6.9, 4.2]
GTS[b]: T_i	Non-Parametric	3.04 1.98-4.71	5.27 3.18-8.30	2.23 (0.65)	92.6 [1.2, 6.2]
	Bootstrap: Percentile	3.06 1.96-4.73	5.22 3.14-8.30	2.15 (0.64)	91.7 [1.4, 6.9]
	Bootstrap: Bias-corrected	2.98 1.93-4.71	5.04 3.14-7.71	2.06 (0.62)	93.0 [2.6, 4.4]
	Standard Asymptotic	3.68 2.55-5.05	5.92 3.98-8.30	2.23 (0.48)	65.4 [0, 34.6]
PA: T_i	Non-Parametric	3.41 2.82-3.95	4.09 3.54-4.75	0.67 (0.17)	92.9 [3.1, 3.9]
	Bootstrap: Percentile	3.44 2.87-3.98	4.06 3.52-4.73	0.62 (0.16)	90.4 [4.5, 5.0]
	Bootstrap:	3.42	4.03	0.61	90.5

	Bias-corrected	2.82 -3.96	3.50-4.72	(0.16)	[5.4, 4.0]
	Standard Asymptotic	3.55 3.08-4.03	4.07 3.54-4.62	0.52 (0.08)	89.1 [6.9, 4.0]
LAP^c: T_i	**Non-Parametric**	3.30 2.80-3.94	4.05 3.51-4.78	0.75 (0.19)	94.5 [4.3, 1.2]
	Bootstrap: Percentile	3.32 2.80-3.94	4.03 3.44-4.72	0.70 (0.18)	92.9 [5.4, 1.7]
	Bootstrap: Bias-corrected	3.29 2.73-3.90	3.99 3.35-4.71	0.70 (0.18)	91.7 [7.3, 1.0]
	Standard Asymptotic	3.40 2.78-3.93	4.08 3.49-5.48	0.68 (0.26)	89.1 [10.6, 0.3]
GTS: Tp	**First Order Delta Method**	2.91 1.26-4.01	4.21 3.04-11.13	1.31 (1.13)	82.2 [17.2, 0.6]
	Second Order Delta Method	2.83 0.11-3.98	4.29 3.09-12.70	1.46 (1.48)	57.4 [42.4, 0.2]
PA: Tp	**First Order Delta Method**	3.45 2.92-3.91	4.04 3.54-4.54	0.59 (0.06)	93.4 [3.3, 3.2]
	Second Order Delta Method	3.41 2.89-3.87	4.07 3.57-4.58	0.66 (0.06)	85.0 [15.0, 0]

[a] Not all Sim3 subjects converged in the GTSS method, although approximately 80% of the trials had PK parameter estimates from 35 or more subjects.

[b] Not all Sim3 subjects converged in the GTS method, although approximately 85% of the trials had PK parameter estimates from 35 or more subjects.

[c] Not all Sim3 subjects converged in the LAP method, although all of the trials had PK parameter estimates from 35 or more subjects.

Discussion

In this chapter we have compared single-sample Tmax confidence interval techniques based on several standard approaches and developed an equation using the second order delta method approach. In addition, we have compared the accuracy and precision of three procedures, non-compartmental, global two-stage, and population-averaged numerical approximation using the Laplace method, with simulated urine excretion data.

Consistent with previous research described in Chapter 3, non-compartmental determination of Ti for urine excretion data led to over-estimation of $\mathbf{Tmax_p}$ and should not be used with urine excretion data. Of the nonlinear methods, the population-averaged Laplace numerical approximation method which is based on maximum likelihood was more accurate and more precise than the global two-stage approach which is based on weighted least-squares for both Ti and Tp. Other advantages of the population-averaged Laplace numerical approximation approach over the global two-stage approach are: (1) it can be used with sparse data; (2) it is much less computationally intensive; and (3) standard programs exist and are easier for the user to adapt to different PK models. The GTS method incorrectly estimated parameter variances, with the highest MSE, bias, and variances resulting from this method.

When enough sampling times within individuals exist, such that estimation of individual subject-level parameters, e.g. Km_i and K_i, is possible, methods based on the median Ti performed better than the asymptotic method based on the mean Ti. Of these median Ti methods, the non-parametric confidence interval method performed the best and was most likely to provide the nominal coverage. The non-parametric method was easy to use and should be considered when calculations of confidence intervals for

median *Ti* are desired. The bootstrap percentile and bias-corrected methods were moderately simple to use and also provided good coverage for median *Ti*. The asymptotic approach performed the worst and was least likely of all the methods to provide a confidence interval that actually included E(**Tmax$_i$**).

When insufficient subject-level data are available, population-average methods can be used to estimate population PK parameters and directly calculate *Tp*. Population-average methods can also be used when sufficient subject-level data are present for generation of both *Ti* and *Tp*, allowing for generation of CIs using any method.   In this application, CIs using the first- and second-order delta method were derived. The first-order delta method provided higher coverages than the second-order method, however never reached the nominal level of coverage. The lower coverages associated with the second-order delta method may have been due to the differences in the expected value for *Tp* calculated as part of the second-order method. While with the first-order delta method E(*Tp*)≈g(**μ**)=**Tmax$_p$**,   when   using   the   second-order   delta   method,

$$E\left(Tp\right) \approx g(\boldsymbol{\mu}) + \frac{1}{2}\left\{ \frac{\partial^2 g(\boldsymbol{\mu})}{\partial^2 \kappa} \times \sigma^2_{K^*_p} + \frac{\partial^2 g(\boldsymbol{\mu})}{\partial^2 \kappa m} \times \sigma^2_{Km^*_p} \right\} \neq \mathbf{Tmax_p.}$$ The   second-order   delta

method may be more appropriate when the function of interest, here g(**v**), is nonlinear over the expected parameter range [Cooch & White 2009]. In our sample, however, the distribution of *Tp* was well-centered (Figure 4.2) when calculated from the PA-method parameter estimates, and had a right skew when calculated from the GTS-method estimates. Figure 4.6 below illustrates the effect of the second-order delta method calculation of E(*Tp*) on the coverages.   In this figure, the 1000 simulated trials were sorted by E(*Tp*) and plotted; E(*Tp*) appears in red, the confidence intervals are represented by the gray horizontal lines and **Tmax$_p$** is represented by the vertical black

hatched reference line. While **Tmax$_p$** appears to be well-centered within the confidence interval ranges, E(*Tp*) is actually over-estimating the center of the confidence intervals, which would result in lower coverages. Without the bias-correction factor consisting of the second-order terms, the second-order delta method coverages would range from 95.9% to 96.1%.
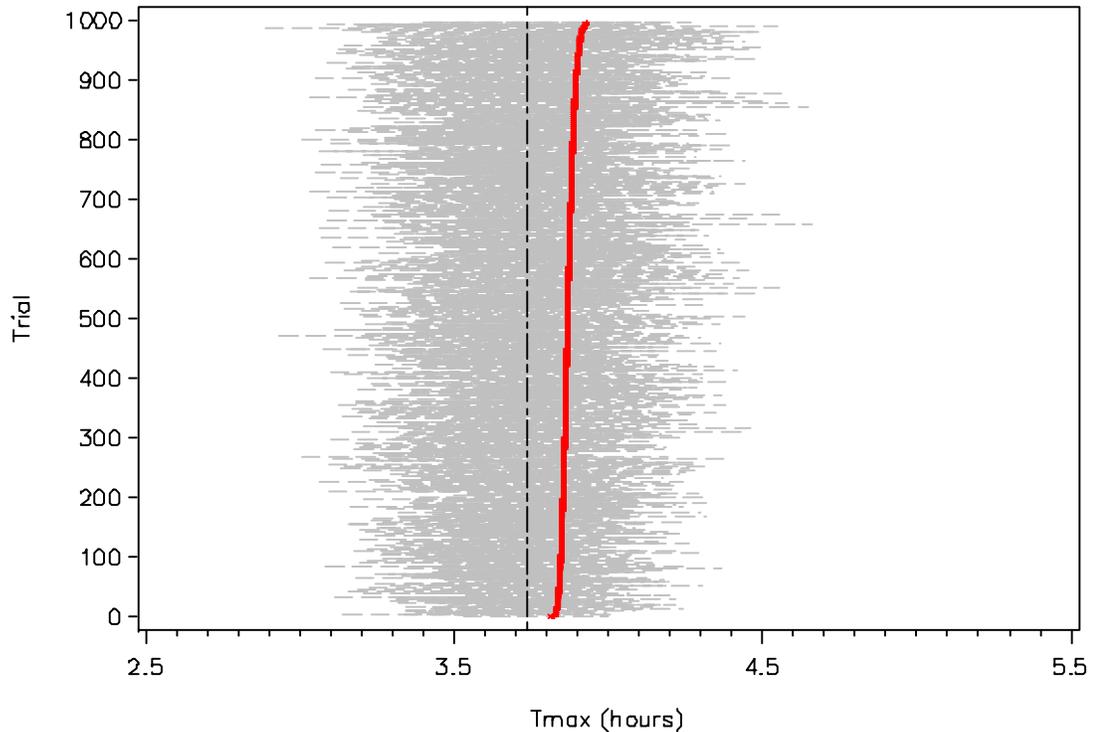
Figure 4.6: Illustration of Second-Order Delta Method Results for the Sim1 data utlilizing the PA-method. Gray horizontal lines represent individual trial confidence intervals, the red line represents E(*Tp*) and the vertical black reference line indicates Tmax$_p$ of 3.74 hours.

The CI approaches considered here included non-parametric, bootstrap percentile, bootstrap bias-corrected, and standard asymptotic calculated using *Ti*; as well as first- and

second-order delta method calculated for Tp when Tp was directly calculated from the nonlinear model estimates. Based on the coverage values for each of the 3 simulations, the non-parametric method performed the best among CIs calculated for Ti, and the first-order delta method performed the best among CIs calculated for Tp. From this research, we recommend that the confidence interval for Tmax should utilize the first-order delta method in conjunction with population-average modeling techniques. If modeling of the population as a whole is not possible and only individual Ti's are calculable, then the non-parametric CI would be recommended.

Returning to the observed 1-AP urine excretion data collected in the DE experiment, we recall that due to the large variety of response profiles among subjects, no single population-average model was identified which would describe all the subjects, therefore only the Ti values were calculated. In this data, the median Ti was 4.85 hours and the non-parametric CI was 2.99 to 5.60 hours. While this CI tells us where the expected center of the Ti distribution is, researchers may also want to know where the range of expected Ti values might occur in order to plan data collection times in future experiments, for example, "How long after an exposure should samples be collected in order to assure 90% probability of collecting at least 80% of the Ti?". Using the non-parametric approach, we could construct a 90% CI around the 80th data percentile to answer this question. For example, for the observed 1-AP data that provided PK parameter estimates, the 80th percentile was 7.3 hours with 90% CI of 5.5 to 16.4 hours, thus indicating that to have 90% assurance of capturing at least 80% of the peak concentration times, one needs to sample until at least between 5.5 and 16.4 hours after the end of the exposure period. As noted earlier, recall that by treating the T_i as **Tmax$_i$**

for quantile estimation, some measurement error may be incorporated and result in biased estimates. While research has shown that attempts to mitigate this effect are ineffective when central quantiles are used, it may be useful to consider these published methods when calculating quantiles away from the center of the distribution [Schechtman and Spiegelman 2007]. Future work should consider this. If we had been able to calculate Tp only, we would consider an asymptotic approach using a tolerance interval.

Appendix A

An overview of the theory for nonlinear mixed effects models is presented below along with a derivation of the likelihood equation. In concert with the notation used by Davidian and Giltinan (2003), a form for the basic model used in this paper is outlined below.

Let y_{ij} denote the jth measurement of the response at time t_{ij}, $j=1,...,n_i$. Letting $\mathbf{y}_i = (y_{i1},...,y_{in_i})^T$, it is usually assumed that the $\mathbf{y}_i$ are independent across i, such that subjects are considered unrelated. The non-linear mixed effects model is then specified as a two-stage hierarchical model, with stage 1 reflecting the individual-level model and stage 2 specifying the population-level parameters. At stage 1,

$$y_{ij} = g(t_{ij}, \beta_i) + \varepsilon_{ij} \qquad (B.1)$$

where g is a nonlinear function dependent on a (px1) vector of parameters β_i specific to subject i and the intra-individual deviations, $\varepsilon_{ij} = y_{ij} - g(t_{ij}, \beta_i)$ are assumed to satisfy $E(\varepsilon_{ij}|\beta_i) = 0$ for all j with variance σ^2 such that $\varepsilon_{ij} \sim N(0, \sigma^2)$. The assumption that the within-group errors ε_{ij} are independent and homoscedastic can be relaxed, but is presented as such here for simplicity.

A more general form of the model incorporates a (vx1) vector of subject-level predictors, $\mathbf{x}_i$, such as weight, gender, and/or creatinine clearance, for example. Here, the ($\mathbf{y}_i$, $\mathbf{x}_i$) are assumed independent across i and the stage one model is written as $y_{ij} = g(t_{ij}, x_i, \beta_i) + \varepsilon_{ij}$. For details and examples of nonlinear models using this more general form with subject-specific covariates, the reader is referred to the following excellent sources [Davidian and Giltinan 1995, Pinheiro and Bates 2004]

At stage 2, a common practice is to apply a linear relationship between β_i and the fixed and random effects, for example,

$$\beta_i = A_i \beta + B_i b_i \qquad \text{(B.2)}$$

where β is the p-dimensional vector of fixed population parameters, b_i is a k-dimensional random effects vector associated with the ith subject and not varying with j, A_i and B_i are design matrices for the fixed and random effects, respectively. The stage 2 modeling characterizes how elements of β_i vary between individual subjects, due to random variation in the population of individuals, denoted by b_i. As a standard assumption, $b_i \sim N(0, D)$, where D is an unstructured covariance matrix that is the same for all individuals, and characterizes the magnitude of unexplained variation in the elements of β_i, and the associations among them.

Specifically, at stage one for the 1-AP urine concentration model, we have:

$$y_{ij} = g\left(t_{ij} + \beta_i\right) + \varepsilon_{ij}$$

$$= \frac{e^{\beta_{2i}} e^{\beta_{3i}} e^{\beta_{4i}} e^{\beta_{5i}}}{e^{\beta_{2i}} - e^{\beta_{1i}}} \left(\frac{1}{e^{\beta_{1i}}} - \frac{1}{e^{\beta_{2i}}} \right) + \qquad \text{(B.4)}$$

$$\frac{e^{\beta_{2i}} e^{\beta_{3i}} e^{\beta_{4i}} e^{D_i}}{e^{\beta_{2i}} - e^{\beta_{1i}}} \left[\frac{1}{e^{\beta_{2i}}} \exp\left(-e^{\beta_{2i}} t_{ij}\right) - \frac{1}{e^{\beta_{1i}}} \exp\left(-e^{\beta_{1i}} t_{ij}\right) \right] + \varepsilon_{ij}$$

where $\beta_{1i}=\log(K_i)$, $\beta_{2i}=\log(Km_i)$, $\beta_{3i}=\log(Ke_i)$, $\beta_{4i}=\log(F_i)$, and $\beta_{5i}=\log(D_i)$, but where D_i =1 in the 1-AP modeling allowing us to drop the $e^{\beta_{5i}}$ term from the model.

At stage two, $\boldsymbol{\beta}_i = \mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i$ which may be written as,

$$\underbrace{\begin{bmatrix} \beta_{1i} \\ \beta_{2i} \\ \beta_{3i} \\ \beta_{4i} \end{bmatrix}}_{\beta_i} = \underbrace{\begin{bmatrix} 1000 \\ 0100 \\ 0010 \\ 0001 \end{bmatrix}}_{A_i} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} 1000 \\ 0100 \\ 0000 \\ 0000 \end{bmatrix}}_{B_i} \underbrace{\begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix}}_{b_i}$$

to show the design matrices, A_i and B_i. In this case, random effects were modeled on only two of the PK parameters, the metabolism rate, *Km*, and the excretion rate, *K*. Further,

$$b_i \sim N\left(0, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & 0 & 0 \\ \Sigma_{21} & \Sigma_{22} & 0 & 0 \\ 0 & 0 & \Sigma_{33} & 0 \\ 0 & 0 & 0 & \Sigma_{44} \end{bmatrix}\right), \quad \varepsilon_{ij} \sim N\left(0, \sigma^2\right)$$

showing that the Σ matrix for the random effects was block diagonal with covariance modeled between the metabolism and excretion rate parameters.

Maximum Likelihood Estimation

The likelihood function for a model is the probability density for the data given the parameters of the model. It is a function which makes the observed data most probable. For a given set of independent, identically distributed observations, $y = (y_1, ..., y_m)$, with density function $f(y_i; \theta)$ for each observation, where θ represents the parameters of interest, in this case $\theta = (\beta, \sigma^2, D)$, the full likelihood is defined as

$$L(\theta \mid y) = \prod_{i=1}^{m} f(y_i; \theta) \tag{B.5}$$

Maximum likelihood estimation in mixed-effects models is more complicated due to the presence of unobserved random effects. Because of these random effects, the

contribution of individual i to the likelihood function is based on the marginal density of the response y_i, calculated as

$$f(y_i \mid \beta, \sigma^2, D) = \int f(y_i \mid b_i, \beta, \sigma^2) f(b_i \mid D) db_i \qquad \text{(B.6)}$$

where $f(y_i \mid \beta, \sigma^2, D)$ is the marginal density of y_i, $f(y_i \mid b_i, \beta, \sigma^2)$ is the conditional density of y_i given the random effects b_i, and $f(b_i \mid D)$ is the marginal distribution of b_i.

Recall that

$$y_{ij} = g(t_{ij}, \beta_i) + \varepsilon_{ij} \qquad i = 1, \ldots, m, \qquad j = 1, \ldots, n_i \qquad \text{(B.7)}$$

where m is the number of subjects and n_i is the number of observations within a subject, g is a general, real-valued, differentiable function of a subject specific parameter vector β_i and covariate vector (t_{ij}), and ε_{ij} is a normally distributed within-subject error term. The function, g is non-linear in at least one component of the β_i, which are modeled as

$$\beta_i = A_i \beta + B_i b_i \quad \text{with } b_i \sim N(0, D) \qquad \text{(B.8)}$$

and it assumed that observations corresponding to different subjects are independent and that the within-subject errors, ε_{ij}, are independent of the random effects, b_i.

The conditional density of the y_i is multivariate normal and can be expressed as

$$f(y_i \mid b_i, \beta, \sigma^2) = \frac{\exp\left(-\sum_{j=1}^{n_i} [y_{ij} - g(t_{ij}, \beta_i)]^2 / 2\sigma^2\right)}{(2\pi\sigma^2)^{n_i/2}} \qquad \text{(B.9)}$$

The distribution of the random effects vectors b_i is completely characterized by its variance-covariance matrix, D, because it is assumed to be normally distributed with mean 0. Therefore, the marginal density of b_i, which is multivariate normal, can be expressed as

$$f\left(b_i \mid D, \sigma^2\right) = \frac{\exp\left(-b_i^T D^{-1} b_i\right)}{(2\pi)^{k/2} \sqrt{|D|}}$$

(B.10)

where we recall that k is the dimension of the random effects vector.

Combining equations (B.9) and (B.10) across all individuals forms the likelihood for a non-linear mixed model defined in (B.5):

$$L\left(\beta, \sigma^2, D \mid y\right) = (2\pi)^{-(N+mk)/2} \left|\sigma^2\right|^{-N/2} |D|^{-m/2} \prod_{i=1}^{m} \int \exp\left\{\frac{\left\|y_i - g_i(\beta, b_i)\right\|^2}{-2\sigma^2} - \left(\frac{1}{2}\right) b_i^T D^{-1} b_i\right\} db_i$$

(B.11)

where $g_i(\beta, b_i) = g(t_{ij}, \beta_i)$ and $N = \sum_{i=1}^{m} n_i$.

Several different statistical techniques for calculating the maximum likelihood of nonlinear mixed effects models have been proposed over the years. These various techniques fall into two basic categories, the two-stage approach and the numerical approximation approach. The two-stage approach is based on first estimating individual subject parameters, i.e. β_i, and then using these values to obtain estimates for β and Σ. According to methodology developed by Davidian and Giltinan, (1993, 1995) the β_i estimates can be obtained using a generalized least-squares / psuedo-likelihood (GLS-PL) algorithm and in the second stage an EM algorithm may be used or, if the user is willing

to assume a linear form for the second stage model, a linear mixed model solution using standard software can be applied. Please see the references for details. Potential drawbacks of this method include the need for sufficiently large n_i on each of the m subjects to make estimation of the β_i feasible, and no standard software is currently in existence for the implementation of this method.

In some instances, data is too sparse to allow for individual estimation of parameters. Many numerical approximation methods have been proposed over the years [Davidian and Giltinan 2003, Sheiner and Beal 1990, Vonesh 1996, Wolfinger 1993, Pinheiro and Bates 1995 and 2004, to list a few] to allow for estimation of the parameters at the population-level, a brief introduction of several of these approaches is provided here. The earliest methods take an approach in which the integral in (B.11) is approximated by a closed form expression in order to maximize an objective function. For example, first order linearization methods are based on an approximation technique built on a first-order Taylor-series expansion with the random effects centered around the expected value of zero (i.e. centered around b_i=0). This is the strategy utilized in the nonmem software package and is generally attributed to Sheiner and Beal (1980). A more refined approach based on first-order conditional methods performs a first-order Taylor series expansion around the conditional modes of the random effects [Lindstrom & Bates, 1990] with alternative derivations of this technique using a Laplace approximation discussed by Wolfinger (1993) and Vonesh (1996). Later methods based on the exact likelihood using adaptive Gaussian quadrature have been developed by Pinheiro and Bates (1995). Many of these techniques are available in standard software packages such as SAS and SPlus. EM algorithm and Bayesian approaches have also

been proposed, with some implementation available in software packages. A comparison of several likelihood approximations presented by Pinheiro and Bates (1995) through evaluation of real data examples and simulations showed that Laplacian and adaptive Gaussian quadrature approximations appeared to provide the best mix of efficiency and accuracy. In this chapter, the Laplacian technique was used.

Appendix B

Three sets of simulated data consisting of 1000 trials with 40 subjects each (as in the evaluable subset of the DE study) were generated to evaluate the CI methods. In the first step of the simulation process, time points representing urine void times were generated for each subject. Under the fixed sampling paradigm used in Sim3, excretion values were generated to mimic a void every 2 hours from 0 to 24 hours post exposure for a total of 13 samples for each subject. When spot sampling was used as in Sim1 and Sim2, subjects were simulated with 4 to 8 urine samples per subject as in the observed data in the DE experiment. The data were generated by first randomly determining the number of time points, n_i, for subject i for each trial. Let X_r $(r = 1,...,5)$ denote the number of subjects with n_i urine samples where n_i was between 4 and 8, then the number of subjects with n_i urine samples followed a multinomial distribution where $X_r \sim Multinomial(0.2, 0.2, 0.2, 0.2, 0.2; 40)$. Spot sampling was mimicked by assuming collection times were randomly and uniformly distributed throughout specified periods according to the following distribution:

Table A.1: Spot Sampling Time Distribution

Sample Number	Time Value Simulation Model
1	Uniform(0, 2)
>1 – max(int(n_i/2))	Uniform(2, 10)
>max(int(n_i/2)) - n_i-1	Uniform(10, 20)
n_i	Uniform(20, 24)

Specifically, each simulated subject provided at least one sample from 0 to 2 hours and another from 20 to 24 hours; the remaining samples were split between the periods from 2 to 10 hours and from 10 to 20 hours post exposure.

Cumulative urine concentration values were generated using the cumulative distribution equation:

$$D_u(t) = \frac{FKeKmD}{Km-K}\left(\frac{1}{K}-\frac{1}{Km}\right) + \frac{FKeKmD}{Km-K}\left[\frac{1}{Km}e^{-Kmt} - \frac{1}{K}e^{-Kt}\right] \qquad (A.1)$$

in a two-stage process. In the first stage, parameter values were determined for each subject. Parameters followed a multivariate normal distribution specified as:

$$\begin{bmatrix} \log(K) \\ \log(Km) \\ \log(Ke) \\ \log(F) \end{bmatrix} \sim MVN\left(\begin{bmatrix} \kappa \\ \kappa m \\ \kappa e \\ \phi \end{bmatrix}, \begin{bmatrix} \Sigma_{K,K} & \Sigma_{K,Km} & 0 & 0 \\ \Sigma_{K,Km} & \Sigma_{Km,Km} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\right)$$

where we note that parameters were exponentiated when modeled to assure positive estimates. In this parameter distribution, *Ke* and *F* were left as fixed values with no associated variance in order to focus on the parameters associated with Tmax calculation, *K* and *Km*. The expected values for log(*K*), log(*Km*), log(*Ke*) and log(*F*) are provided in Table A.2 below.

Table A.2: Expected Parameter Values

Planned Tmax (hours)	Parameter			
	κ	κm	κe	ϕ
3.74	-2.1	-0.7	0.6	3.0

Random effects matrices for log(*K)* and log(*Km)* were determined by setting $\Sigma_{K,K}$=0.03 and $\Sigma_{Km,Km}$=0.12. The correlation, $\rho_{K,km}$, was set to 0.04 in Sim1 and Sim3 and 0.20 in Sim2, such that $\Sigma_{K,Km} = \rho_{K,Km} \times \Sigma_{K,K}^{0.5} \times \Sigma_{Km,Km}^{0.5}$; where $\Sigma_{K,Km}$ is the covariance of the

random effects for log(K) and log(Km). The values for random effects were based on the experimental data results, the correlation in Sim1 and Sim3 was also based on the experimental results; however, the higher correlation for Sim2 was not. As with the experimental data, D, the amount of DE exposed to the subject, was not included in the modeling as this was standardized as part of the experiment.

In stage two, y_{ij}, the cumulative urine concentration for the ith subject ($i=1,\ldots, m$) at the jth time point, ($j=1,\ldots, n_i$) was calculated as $y_{ij}=Du(t_{ij})+\varepsilon_{ij}$ where the ε_{ij} are the within-subject errors and $\varepsilon_{ij} \overset{iid}{\sim} N(0,21)$. The variance of the ε_{ij} was set using the observed 1-AP excretion data. This resulted in three simulated data sets of cumulative urine concentration values with random within- and between-subject variability for Tmax.

Appendix C

The computational details on the derivation of σ_{Tp}^2, the Var[g(**v**)] using the

second order Delta Method are provided below.   Recall that Tp, the population-averaged

time of maximum concentration may be calculated by:

$$g(\mathbf{v}) = (Km\text{-}K) / \exp(Km)\text{-}\exp(K)$$

where, for simplification we have dropped the "*" and "p" notations used to denote the

transformed population-averaged values.

First, we state the general form for a second order Taylor Series expansion of a

function of two variables, $g(\mathbf{v})$, $\mathbf{v} = (K, Km)^T$, about the point $\boldsymbol{\mu} = (\kappa, \kappa m)^T$:

$$g(\mathbf{v}) \approx g(\boldsymbol{\mu}) +$$
$$\frac{\partial g(\boldsymbol{\mu})}{\partial \kappa} \times (K - \kappa) + \frac{\partial g(\boldsymbol{\mu})}{\partial \kappa m} \times (Km - \kappa m) + \qquad \text{C.1}$$
$$\frac{1}{2} \left\{ \begin{array}{l} \dfrac{\partial^2 g(\boldsymbol{\mu})}{\partial^2 \kappa} \times (K - \kappa)^2 + 2 \times \dfrac{\partial^2 g(\boldsymbol{\mu})}{\partial \kappa \partial \kappa m} (K - \kappa)(Km - \kappa m) + \\ \dfrac{\partial^2 g(\boldsymbol{\mu})}{\partial^2 \kappa m} \times (Km - \kappa)^2 \end{array} \right\}$$

where the subscripts are used to denote the respective partial derivatives.

Second, we consider a general formula for the central moments of order k based on a

paper by Triantafyllopoulos (2003) that we will use in our calculation of Var[g(**v**)].

Given $V = (V_1, ..., V_k)'$ follows a normal distribution with known mean ξ and variance

$C = \{c_{ij}\}$, $i,j = 1, ..., k$, the central moments, $\mu_{l,...,k}(V\text{-}\xi)$ may be calculated such that

(a) if k is odd, $\mu_{l,...,k}(V\text{-}\xi) = 0$, and
(b) if k is even with $k = 2\lambda$ ($\lambda \geq 1$), then $\mu_{l,...,k}(V\text{-}\xi) = \Sigma(c_{ij}c_{kl}, ..., c_{xz})$, where the sum is taken over all the permutations of $\{1, ..., 2\lambda\}$ giving $(2\lambda\text{-}1)!/(2^{\lambda\text{-}1}(\lambda\text{-}1)!)$ terms in the sum, each a product of λ covariances.

For example, four formulas we will use in our variance derivation are:

(1) $E_c(K^4) = 3\sigma_K^2$;

(2) $E_c(K^3 Km) = 3\sigma_K^2 \sigma_{K,Km}$

(3) $E_c(K^2 Km^2) = \sigma_K^2 \sigma_{Km}^2 + 2(\sigma_{K,Km})^2$; and

(4) $E_c(KKm) = \sigma_{K,Km}$

where E_c denote the expectation under the central moment theory, i.e. $E_c(K^4) = E((K-\kappa)^4)$ and $E_c(K^2 Km^2) = E[(K-\kappa)^2(Km-\kappa m)^2]$ but $E_c(KKm^2) = 0$.

To determine the variance, we re-write equation C.1 as follows, showing all variance and covariance terms, where we have simplified the notation by letting A=

$\dfrac{\partial g(\mu)}{\partial \kappa}$, B= $\dfrac{\partial g(\mu)}{\partial \kappa m}$, C= $\dfrac{\partial^2 g(\mu)}{\partial^2 \kappa}$, D= $\dfrac{\partial^2 g(\mu)}{\partial^2 \kappa m}$, and F= $\dfrac{\partial^2 g(\mu)}{\partial \kappa \partial \kappa m}$:

$$
\begin{aligned}
Var[g(\mathbf{v})] \approx{}& Var\big[A(K-\kappa) + B(Km-\kappa m) + 0.5C(K-\kappa)^2 + F(K-\kappa)(Km-\kappa m) + 0.5D(Km-\kappa m)^2\big] \\
={}& A^2 Var(K-\kappa) + B^2 Var(Km-\kappa m) + 0.25C^2 Var[(K-\kappa)^2] + F^2 Var[(K-\kappa)(Km-\kappa m)] + \\
& 0.25D^2 Var[(Km-\kappa m)^2] + 2AB \times Cov[(K-\kappa),(Km-\kappa m)] + AC \times Cov[(K-\kappa),(K-\kappa)^2] + \\
& 2AF \times Cov[(K-\kappa),(K-\kappa)(Km-\kappa m)] + AD \times Cov[(K-\kappa),(Km-\kappa m)^2] + \\
& BC \times Cov[(Km-\kappa m),(K-\kappa)^2] + 2BF \times Cov[(Km-\kappa m),(K-\kappa)(Km-\kappa m)] + \\
& BD \times Cov[(Km-\kappa m),(Km-\kappa m)^2] + CF \times Cov[(K-\kappa)^2,(K-\kappa)(Km-\kappa m)] + \\
& 0.5CD \times Cov[(K-\kappa)^2,(Km-\kappa m)^2] + FD \times Cov[(K-\kappa)(Km-\kappa m),(Km-\kappa m)^2]
\end{aligned}
$$

To calculate $Var[g(\mathbf{v})]$, we use the above equation and the central moments defined previously. For example, $Var(K-\kappa)^2 = E[(K-\kappa)^2]^2 + [E(K-\kappa)^2]^2 = E_c[K^2]^2 - [E_c K^2]^2 = E_c K^4 - [E_c K^2]^2$ and $Cov[(K-\kappa),(Km-\kappa m)] = E[(K-\kappa)(Km-\kappa m)] - [E(K-\kappa)E(Km-\kappa m)] = E_c[KKm] - [E_c K][E_c Km] = E_c[KKm]$. The variance formula simplifies to:

$$
\begin{aligned}
Var[g(\mathbf{v})] \approx{}& \sigma_K^2(A^2 + 0.75C^2) + \sigma_{Km}^2(B^2 + 0.75D^2) + (\sigma_{K,Km})^2 \times (F^2 + CD) + F^2 \sigma_K^2 \sigma_{Km}^2 \\
& - 0.25(C^2 \sigma_K^4 + D^2 \sigma_{Km}^4) + 2AB\sigma_{K,Km} + 2CF\sigma_K^2 \sigma_{K,Km} + 2FD\sigma_{Km}^2 \sigma_{K,Km}
\end{aligned}
$$

where we use the parameter estimates provided by the 2-stage or the population average

methods for $\sigma_{K,Km}, \sigma_K^2, and\ \sigma_{Km}^2$, and the first and second partial derivatives for g(κ,κm)

specified below.

First and Second Partial Derivatives of g(κ,κm)

$$\text{Let A} = \frac{\partial g(\mathbf{\mu})}{\partial \kappa} = \frac{\exp(\kappa) - \exp(\kappa m) + \exp(\kappa) \times (\kappa m - \kappa)}{\left[\exp(\kappa m) - \exp(\kappa)\right]^2}$$

$$\text{Let B} = \frac{\partial g(\mathbf{\mu})}{\partial \kappa m} = \frac{\exp(\kappa m) - \exp(\kappa) - \exp(\kappa m) \times (\kappa m - \kappa)}{\left[\exp(\kappa m) - \exp(\kappa)\right]^2}$$

$$\text{Let C} = \frac{\partial^2 g(\mathbf{\mu})}{\partial^2 \kappa} = \frac{\exp(\kappa) \times \{\exp(\kappa m) \times (\kappa m - \kappa - 2) + \exp(\kappa) \times (\kappa m - \kappa + 2)\}}{\left[\exp(\kappa m) - \exp(\kappa)\right]^3}$$

$$\text{Let D} = \frac{\partial^2 g(\mathbf{\mu})}{\partial^2 \kappa m} = \frac{\exp(\kappa m) \times \{\exp(\kappa m) \times (\kappa m - \kappa - 2) + \exp(\kappa) \times (\kappa m - \kappa + 2)\}}{\left[\exp(\kappa m) - \exp(\kappa)\right]^3}$$

$$\text{Let F} = \frac{\partial^2 g(\mathbf{\mu})}{\partial \kappa \partial \kappa m} = \frac{\exp(\kappa m) \times \{\exp(\kappa m) - 2\kappa m \times \exp(\kappa)\} - \exp(\kappa) \times \{\exp(\kappa) - 2\kappa \times \exp(\kappa m)\}}{\left[\exp(\kappa m) - \exp(\kappa)\right]^3}$$

Appendix D
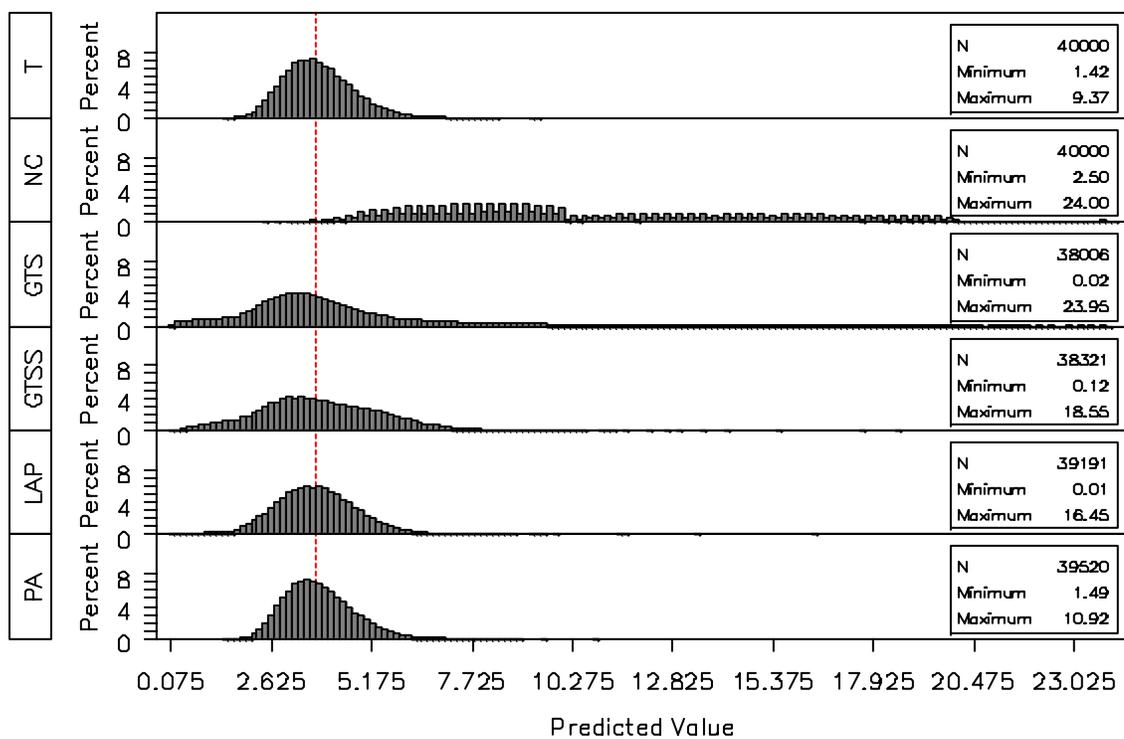
Figure D.1: Distribution of *Ti* Values [rows 2 to 6] as compared to Tmax$_i$ values [row 1], with a reference line indicating Tmax$_p$ of 3.74 hours for the Sim 2 data.
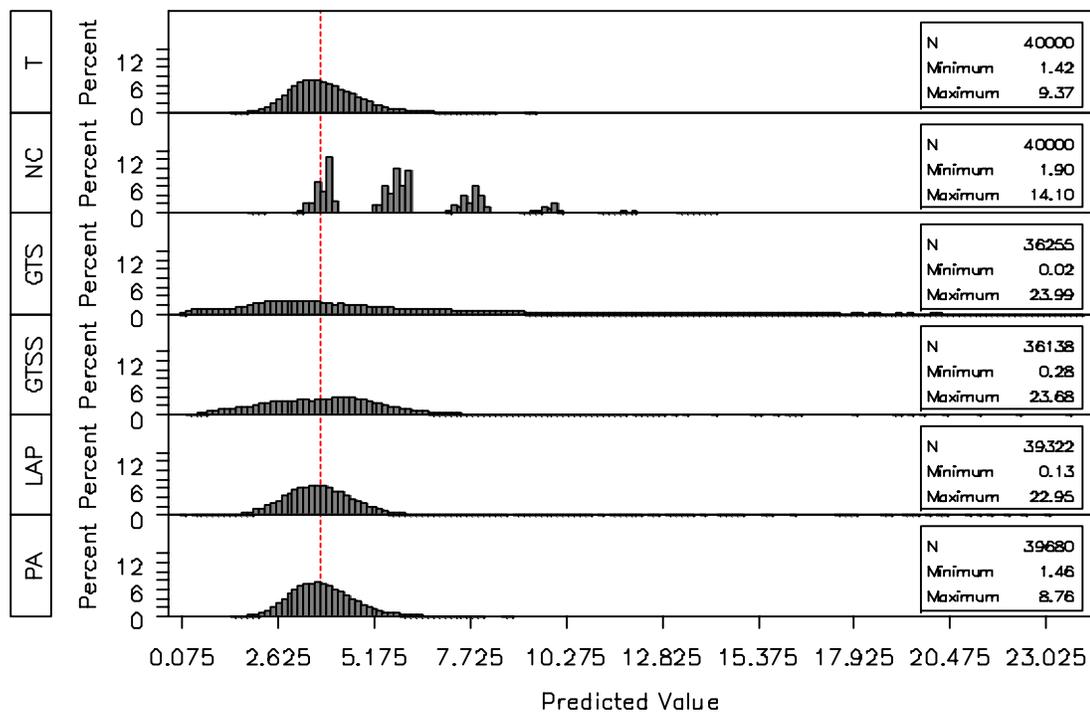
Figure D.2: Distribution of *Ti* Values [rows 2 to 6] as compared to Tmax$_i$ values [row 1], with a reference line indicating Tmax$_p$ of 3.74 hours for the Sim 3 data.
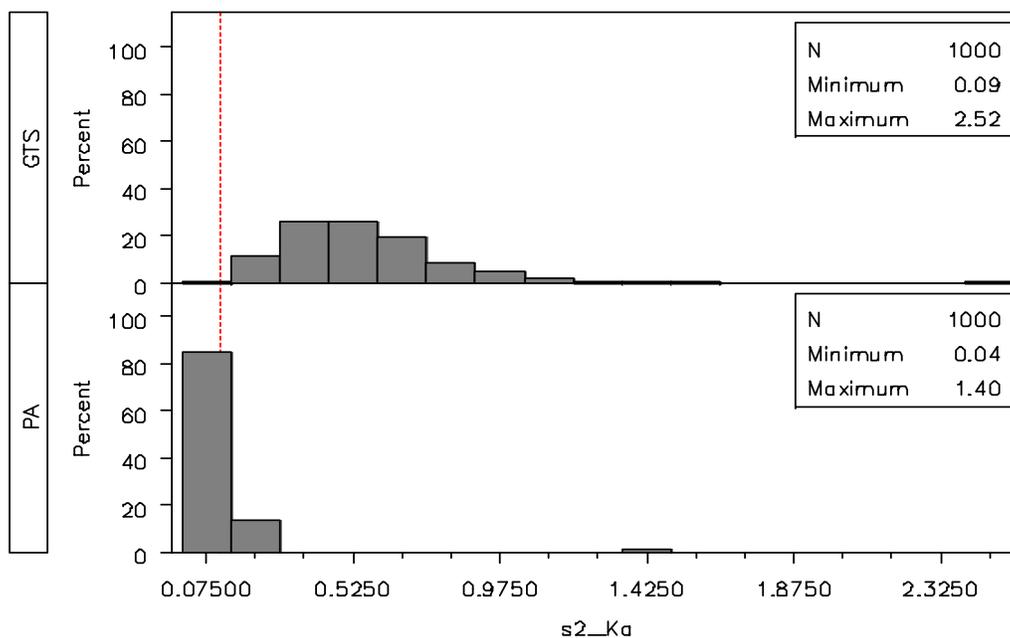
Figure D.3: Sim 2 Σ_{Km} estimated by 2-stage (GTS) and Population-average (PA) nonlinear modeling, reference line indicates "true" value of 0.12.
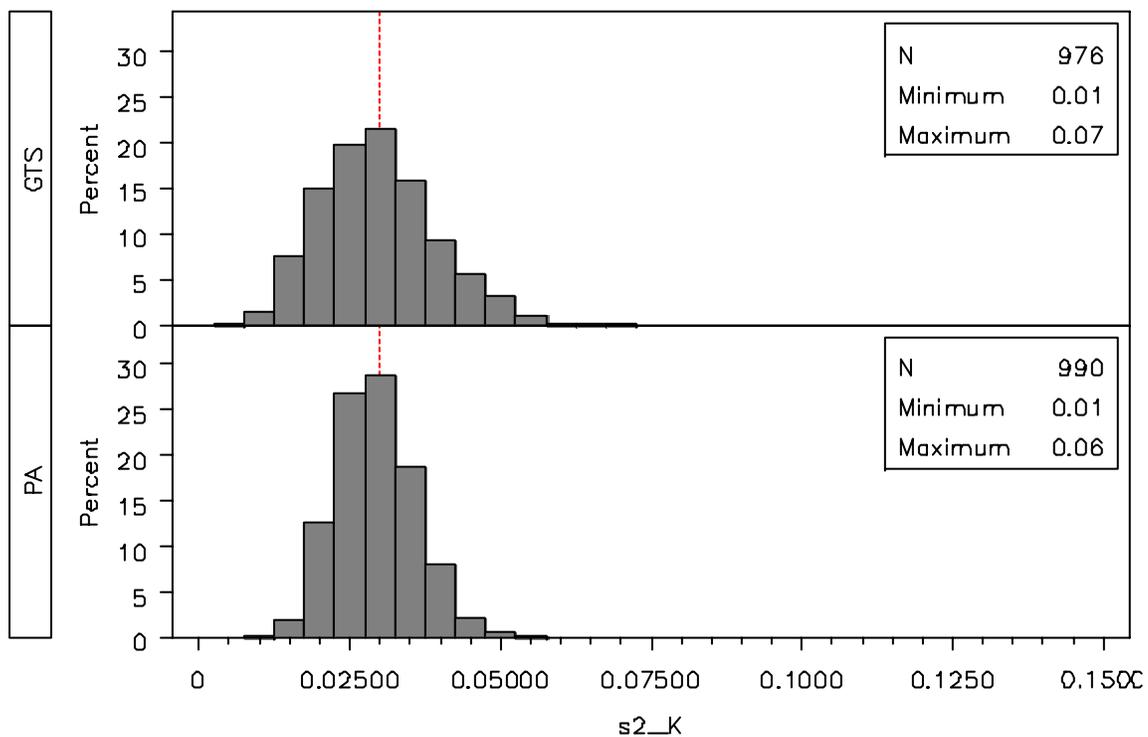
Figure D.4: Sim 2 Σ_K estimated by 2-stage (GTS) and Population-average (PA) nonlinear modeling, reference line indicates "true" value of 0.03.
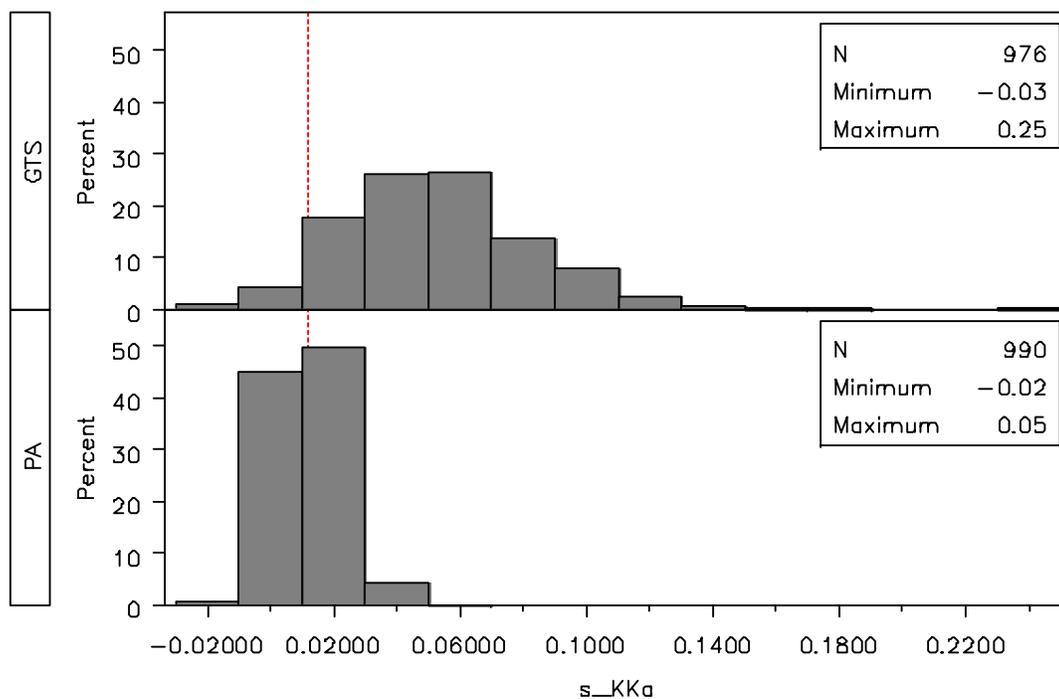
Figure D.5: Sim 2 $\Sigma_{K,Km}$ estimated by 2-stage (GTS) and Population-average (PA) nonlinear modeling, reference line indicates "true" value of 0.012.
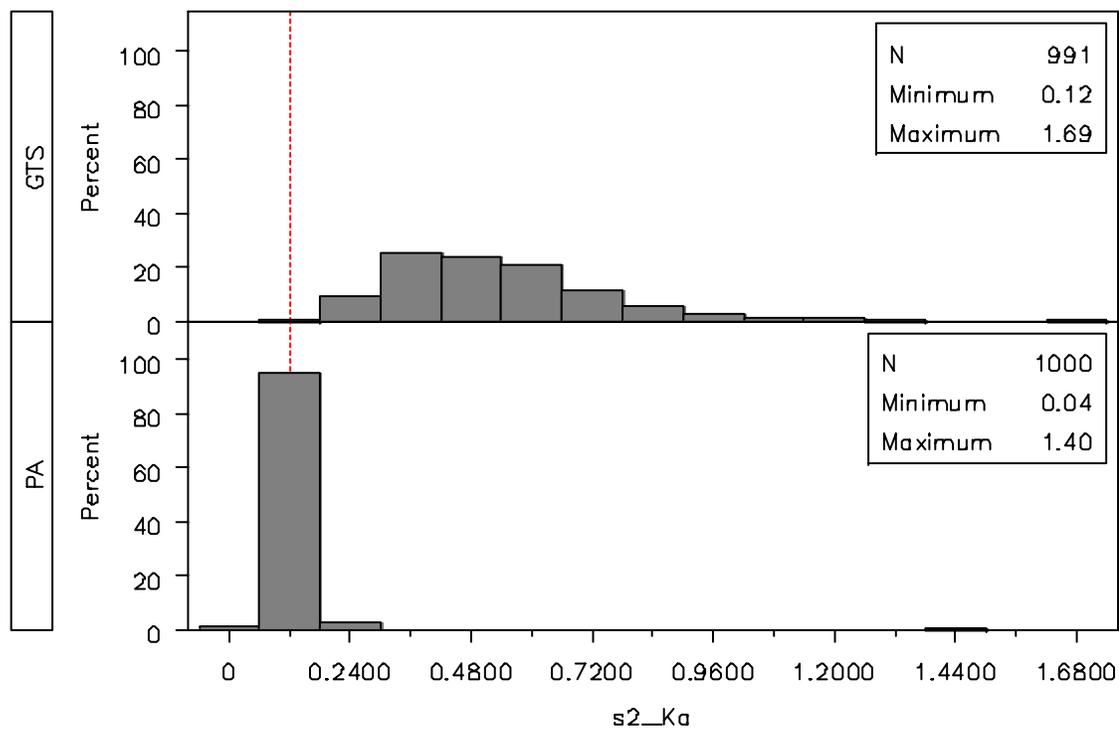
Figure D.6: Sim 3 Σ_{Km} estimated by 2-stage (GTS) and Population-average (PA) nonlinear modeling, reference line indicates "true" value of 0.12.
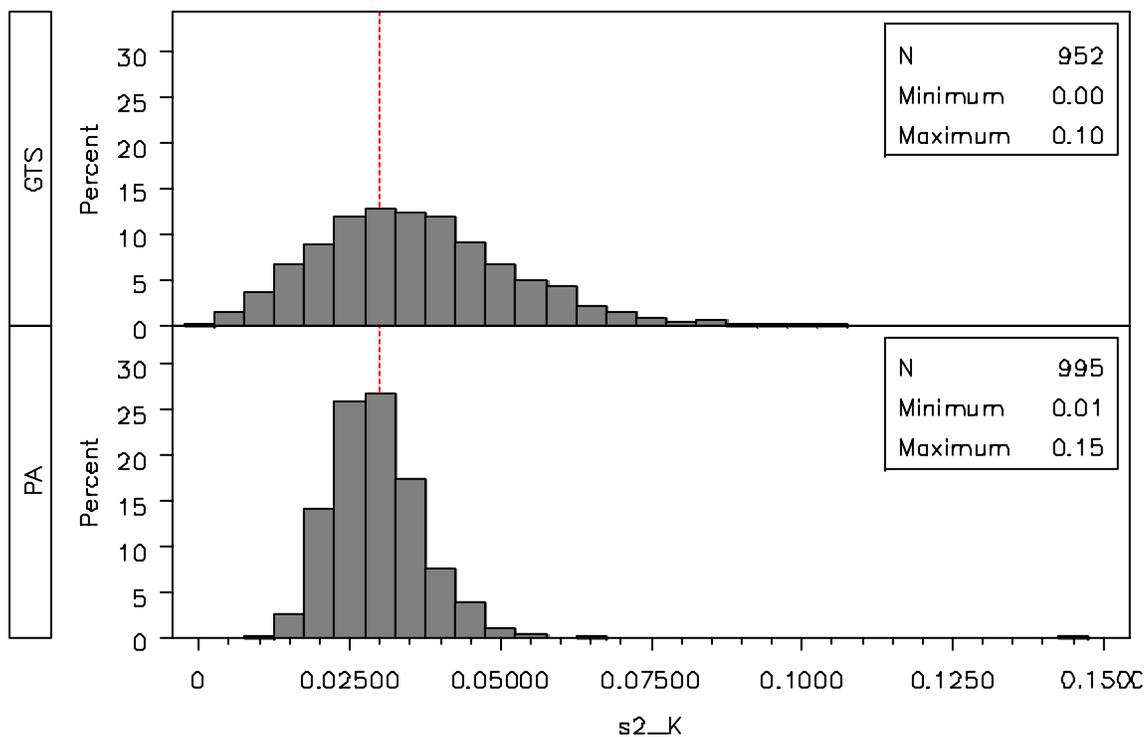
Figure D.7: Sim 3 Σ_K estimated by 2-stage (GTS) and Population-average (PA) nonlinear modeling, reference line indicates "true" value of 0.03.
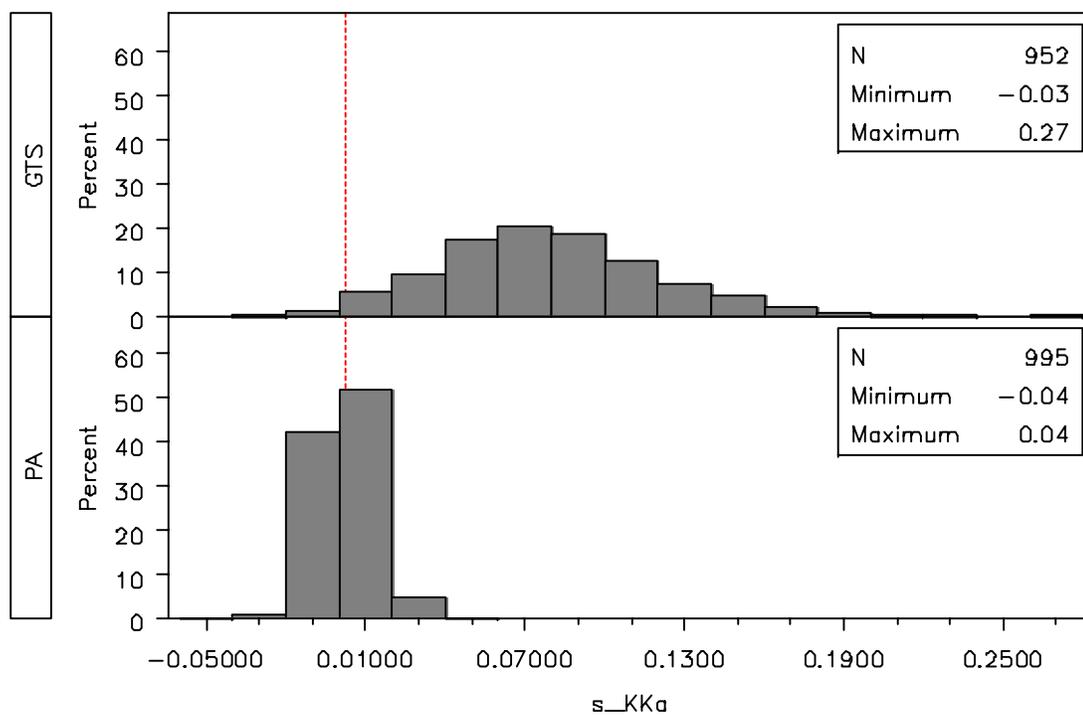
Figure D.8: Sim 3 $\Sigma_{K,Km}$ estimated by 2-stage (GTS) and Population-average (PA) nonlinear modeling, reference line indicates "true" value of 0.0024.

References

Beland, F. A.; Kadlubar, F. F. Metabolic Activation and DNA Adducts of Aromatic Amines and Nitroaromatic Hydrocarbons. In *Carcinogenesis and Mutagenesis I* ; Cooper, C.S.; Grover, P.L., Eds.; Springer-Verlag: New York, 1990, 267-325.

Boos, D. D. & Stefanski, L. A. (2006). *Modern Statistical Inference: Theory and Methods.* New York: Springer.

Cardiello, P. Kroon, E. oeelmans, R., Vorarien, W., Ubolyam, S., Phanuphak, P., Lange, J., Cooper, D., Ruxrungthan, K. (2000). Steady-state pharmacokinetics (PK) of saquinavir (SQV) with and without two doses of itraconazole in HIV-1+ Thai patients:HIV NAT001.2 substudy. Int Conf AIDS, abstract no. WeOrB545.

Charles, B.G., Miller, A.K., Nasveld, P.E., Reid, M.G., Harris, I.E., Edstein, M.D. (2007). Populations Pharmacokinetics of Tafenoquine during Malaria Prophylaxis in Healthy Subjects. *Antimicrobial Agents and Chemotherapy*, 2709-2715.

Chow, S-C & Liu, J-P. (1992). *Design and Analysis of Bioavailability and Bioequivalence Studies.* New York: Marcel Dekker, Inc.

Cooch, E.G. & White, G.C. [online]. 2009. The Delta method…. http://www.phidot.org/software/mark/docs/book/pdf/app_2.pdf (last accessed June 2009).

Cornell, R. G. (1991). Nonparametric Tests of Dispersion for the Two-Period Crossover Design. *Commun. Statist. Theory Meth.* 20(3), 1099-1106.

Davidian, M. & Giltinan, D. M. (1993). Some Simple Methods for Estimating Intra-individual Variability in Nonlinear Mixed Effects Models. *Biometrics,* 49, 59-73.

Davidian, M. & Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data.* Boca Raton: Chapman & Hall.

Davidian, M. & Giltinan, D. M. (2003). Nonlinear Models for Repeated Measurement Data: An Overview and Update. *Journal of Agricultural, Biological & Environmental Statistics, 8* (4) 387-419.

Efron, B. & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science,* 1 (1) 54-77.

Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association,* 82 (397) 171-185.

Goodman, L. A. (1962). The Variance of the Product of K Random Variables. *Journal of the American Statistical Association*, 57(297) 54-60.

Hall, P., (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hauschke, D. Steinijans, V.W., Diletti, E. (1990). A distribution-free procedure for the statistical analysis of bioequivalence studies. *International Journal of Clinical Pharmacology, Therapy and Toxicology*, (28) 72-78.

Hogg, R. V. & Craig, A. T. (1978) *Introduction to Mathematical Statistics*. Macmillan Publishing Co., Inc.: New York.

IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. http://monographs.iarc.fr/ENG/Monographs/vol46/volume46.pdf (accessed July 30, 2008). Volume 46 Diesel and Gasoline Engine Exhausts and Some Nitroarenes.

Islinger, F., Dehghanyar, P., Sauermann, R., Burger, C., Kloft, C., muller, M., Joukhadar, C. (2006). The effect of food on plasma and tissue concentrations of linezolid after multiple doses. *International Journal of Antimicrobial Agents*, 27, 108-112.

Laumbach, R. (2009). Quantification of 1-aminopyrene in human urine after a controlled exposure to diesel exhaust. *Journal of Environmental Monitoring*, (11) 153-159.

Lindstrom, M. & Bates, D. (1990). Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics,* (46) 673-687.

McLachlan, G. J. & Krishnan, T. (1997). *The EM Algorithm and Extensions.* New York: John Wiley & Sons, Inc.

Mortimer, K.J., Tattersfield, A.E., Tang, Y., Wu, K., Hochhaus, S. , Harrison, T.W. (2007). Plasma concentrations of Fluticasone propionate and budesonide following inhalation: effect of induced bronchoconstriction, *British Journal of Clinical Pharmacology*, 64(4), 439-444.

Pinheiro, J. C. & Bates, D. M., (1995). Approximations to the Log-Likelihood function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics,* 4 (1) 12-35.

Pinheiro, J. C. & Bates, D. M., (2004). *Mixed Effects Models in S and S-Plus.* New York: Springer Science+Business Media, LLC.

SAS Institute Inc. *SAS OnlineDoc®,* Version 9.1. SAS Institute Inc.: Cary, NC, 2004.

Schechtman, E., Spiegelman, C. (2007). Mitigating the effect of measurement errors in quantile estimation. *Statistics & Probability Letters*, 77, 514-524.

Shao, J. & Tu, D., (1995). *The Jackknife and Bootstrap.* New York: Springer-Verlag.

Shargel, L. & Yu, A. B. C. (1993). *Applied Biopharmaceutics and Pharmacokinetics.* Stamford: Appleton & Lang.

Sheiner, L. B. & Beal, S. L. (1980). Evaluation of Methods for Estimating Population Pharmacokinetic Parameters. I. Michaelis-Menten Model: Routine Clinical Pharmacokinetic Data. *Journal of Pharmacokinetics and Biopharmaceutics,* 8, F553-571.

Smith, P.F., Forrest, A., Ballow, C.H., Adams, J.M., Shiveley, L.R. (2002). Single dose pharmacokinetics of SPD-756 in healthy adult volunteers, Int. Conf. AIDS, abstract no. WePeB6050.

Tokiwa, H.; Ohnishi, Y. (1986). Mutagenicity and carcinogenicity of nitroarenes and their sources in the environment. *Crit Rev Toxicol.. 17(1),* 23-60.

Triantafyllopoulos K., (2003). On the central moments of the multidimensional Gaussian distribution. The Mathematical Scientist, 18 125-128.

Vonesh, E. F. (1996). A Note on the Use of Laplace's Approximation for Nonlinear Mixed-Effects Models. *Biometrika,* 83, 447-452.

Vonesh, E. & Carter, R. (1992). Mixed-Effects Nonlinear Regression for Unbalanced Repeated Measures. *Biometrics,* (48) 1-17.

Willavize, S.A. & Morgenthien, E. A. (2008). Nonparametric confidence intervals for Tmax in sequence-stratified crossover studies. *Pharmaceutical Statistics*, (7), 9-19.

Wolfinger, R. (1993). Laplace's Approximation for Nonlinear Mixed Models. *Biometrika*, 80, 791-795.

Chapter 5

An EM Algorithm to Estimate Times of Maximum Urine Excretion of Diesel Exhaust Biomarker 1-aminopyrene when a Mixture Distribution is Present

Abstract

Individual differences in the metabolism of 1-nitropyrene, a chemical component of

diesel exhaust, may result in variations in times of maximum excretion of biomarkers

associated with 1-nitropyrene. Previous research led to a hypothesis that a mixture of 2

or 3 distributions describing the times of maximum excretion of 1-aminopyrene, one such

diesel exhaust biomarker, may be present. In this chapter, pharmacokinetic parameters

estimating the metabolism and excretion rates of 1-aminopyrene were used in an EM

algorithm to characterize the hypothesized mixture distribution using a mixture of two

multivariate normal distributions. Results showed that it was difficult to find a separation

in the observed data. Simulated data was used to successfully demonstrate the

application.

Introduction

Exposure to diesel exhaust (DE) represents a public health concern due to the

environmental contaminants contained in DE and DE particles. These environmental

contaminants are nitropolycyclic aromatic hydrocarbons (nitro-PAHs) which are found in

a variety of sources including DE and coal combustion fly ash [Howard, et al. 1995 and

Bond, et al. 1986]. A sub-class of nitro-PAHs, the nitropyrenes, has been shown to have

genotoxic and carcinogenic properties in experimental animals [IARC Monographs 1989,

Tokiwa & Ohnishi 1986, Beland & Kadlubar, 1990]. Concerns about the public health

risks resulting from inhalation exposures to nitropyrenes have inspired numerous studies to further the scientific knowledge on the deposition, metabolism and excretion of nitropyrenes [Bond, et al 1986, Sun, et al. 1983, Dutcher, et al. 1985]. A recent study found that concentrations of nitro-PAHs were 3 orders of magnitude higher in DE particles than in urban ambient particles [Bamford 2003]. The most abundant nitro-PAH in DE particles is 1-nitropyrene (1-NP) which is metabolized to several compounds, one of which is 1-aminopyrene (1-AP).

Urine biomonitoring is a tool that may be used in cross-sectional or cohort studies of environmental-health associations in large populations, for example, exposure to DE or DE particles. Characterization of the trends in metabolite concentration profiles in urine may be important for understanding results and for improving and simplifying studies, as planned urine collection can be made at an optimal time to capture the exposure. Knowledge of the time-course of excretion following DE exposure, presently unknown for 1-AP, would allow for optimization of the sampling times in future studies and more efficient biomonitoring.

The time of maximum excretion (Tmax) is a standard pharmacokinetic (PK) parameter used to characterize the time-course of excretion. In this chapter, we will further our understanding of the Tmax of the DE urine biomarker, 1-AP, following known exposure. We have shown that this Tmax is best estimated using nonlinear PK modeling techniques and compared confidence interval methods for characterizing Tmax in Chapter 4. However, previous work to estimate the time of maximum urine excretion of 1-AP from a recent experiment described in Chapter 3, revealed not only that a wide variety of subject response profiles were present, but hypothesized that this variety of

response patterns may represent a mixture of profiles. Biological explanations to support this hypothesis include potential differences among subjects in metabolic capacities (polymorphisms) [IPCSINTOXDatabank] or differences in exposure based on respiration rates, etc. Therefore, in the case of a mixture of response profiles, as may be present in the DE data, failure to separate the groups may lead to incorrect estimation and characterization of Tmax. In this chapter we seek to further understand the observed 1-AP data by using an Expectation-Maximization (EM) algorithm to identify the components of the hypothesized mixture distribution.

Materials and Methods

DE Experiment

In a recent experiment, spot urine samples representing partial voids were collected from 55 healthy volunteers during a 24-hour period following a brief controlled exposure to diluted diesel exhaust (DE). In the experiment, each study participant underwent two one-hour exposure sessions: one was a diluted DE atmosphere and the other was a filtered clean air atmosphere (CA). The overall goal of the experiment was to examine whether 1-aminopyrene (1-AP) could serve as an adequate biomarker for DE exposure. Published results indicate a significant difference between DE and CA sessions with respect to the average urine concentrations of 1-AP in the 24 hours following exposure [Laumbach et al., 2009]. An additional goal was to characterize the excretion time course of the biomarker 1-AP, in order to optimize sampling times in future studies

Background

We consider an approach based on the results of the observed DE experiment. Although the results of previous analysis in Chapter 3 led to a hypothesis that a mixture of 3 types of response profiles may be present in the observed DE data, lack of sampling longer than the planned 24-hour follow-up period did not allow for modeling of the subset of subjects that may have had maximum excretion times later than 24 hours. Therefore, only the 25 subjects with estimable PK parameters were included in this application of the EM algorithm used to characterize a mixture of two groups. For discussion purposes, we will refer to one group as "responders" and the other group as "non-responders".

In this chapter, we use the following notation for Tmax: define **Tmax1 and Tmax2** as the true population values for the "responder" and "non-responder" groups respectively, Tm_i $(i=1,…, m)$ as random variables which estimate individual Tmax values for the ith subject with E(Tm_i |responder)=**Tmax1** and E(Tm_i |non-responder)=**Tmax2**, $T1$ and $T2$ as random variables which estimate **Tmax1** for "responders" and **Tmax2** for "non-responders", respectively. In this approach, the PK parameters, K_i and Km_i, the renal excretion rate and the rate of 1-NP metabolism to 1-AP for each individual subject were used to calculate Tm_i using the following formula [Shargel and Yu 1993]:

$$Tm_i = \frac{\log(Km_i / K_i)}{(Km_i - K_i)} \qquad (5.1)$$

This estimation may be done using individual subject parameter estimates, K_i and Km_i to calculate Tm_i, or population-averaged estimates, K_p and Km_p to calculate Tm_p, the population-average time of maximum excretion detailed in Chapter 4, depending on the data and modeling approach. For example, with sparse data, population-average estimation methods must often be used because insufficient individual subject-level data

is available for estimation of K_i and Km_i. This is often the case in biomonitoring applications. However, in the analysis of the observed DE data, nonlinear modeling using the Laplace approximation method was used to fit individual subject data due to the wide variety of individual response profiles. To obtain estimates of K_i and Km_i for calculation of Tm_i, K_i and Km_i were reparameterized using logarithms in order to assure positivity of the PK parameter estimates and make fitting of the model to individual data more stable [Chapter 4, Davidian and Giltinan 1995]; denote $Km_i*=\log(Km_i)$; $K_i*=\log(K_i)$. In accordance with the nonlinear mixed modeling theory described in Chapter 4, Appendix B, Km_i* and K_i* were modeled according to a multivariate normal distribution,

$$\begin{bmatrix} \log(K_i) \\ \log(Km_i) \end{bmatrix} = \begin{bmatrix} K_i^* \\ Km_i^* \end{bmatrix} \sim MVN\left[\begin{pmatrix} \kappa \\ \kappa m \end{pmatrix}, \begin{pmatrix} \sigma_{K_p^*}^2 & \sigma_{K_p^*,Km_p^*} \\ \sigma_{K_p^*,Km_p^*} & \sigma_{Km_p^*}^2 \end{pmatrix}\right] \qquad (5.2)$$

Ideally, PK data like the urine excretion results would be modeled using a nonlinear mixed model approach to allow for estimation of the fixed effects as well as the random effects within and between subjects. However, reliable estimation of the within and between subject random effects was not possible in this application due to: (1) the wide variety of response profiles not allowing for estimation of the PK parameters using a single, population-average model; and (2) insufficient subject-level information not allowing for estimation of the $Km_i *$ and $K_i *$ standard errors due to the sparseness of the observed DE data. As a result, the estimated covariance matrix for $Km_i *$ and $K_i *$ will include a combination of within and between subject error, instead of only between-subject error. If we consider that in many PK applications, the within-subject error is typically small compared to the between subject error, then the overestimation of this

covariance matrix should not occlude the true characterization of **Tmax1 and Tmax2** by too much [Davidian and Giltinan 1995 and 2003].

Simulated Data

In addition to using the observed 1-AP urine excretion data, a simulated mixture data set was used to assess the EM algorithm application. The simulated data set consisted of 40 individual subjects divided equally into 2 groups with different expected values but identical covariance matrices, i.e. with **Tmax1**=1.51 hours and **Tmax2** =3.74 hours, specifically:

$$\begin{bmatrix} K_i^* \\ Km_i^* \end{bmatrix} \sim MVN \left[\begin{pmatrix} -0.50 \\ -0.33 \end{pmatrix}, \begin{pmatrix} 0.03 & 0.0024 \\ 0.0024 & 0.12 \end{pmatrix} \right] \text{ when Group=1, and;}$$

$$\begin{bmatrix} K_i^* \\ Km_i^* \end{bmatrix} \sim MVN \left[\begin{pmatrix} -2.10 \\ -0.70 \end{pmatrix}, \begin{pmatrix} 0.03 & 0.0024 \\ 0.0024 & 0.12 \end{pmatrix} \right] \text{ when Group=2.}$$

Data that was simulated for the mixture distribution in Chapter 3 was used in this exercise. In this case, the PK parameter values used to define the "true" individual subject-level Tmax, **Tmax$_i$**, K_i^* and Km_i^*, were used in the EM algorithm for mixture distributions as detailed below.

EM Algorithm: General Theory

The Expectation-Maximization (EM) algorithm [Dempster, Laird, and Rubin 1977] is a broadly applicable approach to the iterative computation of maximum likelihood (ML) estimates in incomplete-data problems. With each iteration of the

algorithm there are two steps, the expectation (E-step) and maximization (M-step). In the E-step the complete log-likelihood is replaced by its conditional expectation, given the observed data. The M-step performs a maximum likelihood estimation of this "completed" data under the current value of the parameters. Starting with some user-provided initial parameter estimates, the E- and M-steps are repeated until convergence.

Specifically, in this context, $\mathbf{Y}$ is observed and $\mathbf{X}$ is unobserved. Assume that the joint density for the complete data $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ is given by:

$$f(\mathbf{Z} \mid \boldsymbol{\theta}) = f(\mathbf{X}, \mathbf{Y} \mid \boldsymbol{\theta}) = f(\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\theta}) f(\mathbf{X} \mid \boldsymbol{\theta})$$

where θ are fixed parameters. Based on this density function, the likelihood function for the complete-data likelihood can be defined as $L(\boldsymbol{\theta} \mid \mathbf{Z}) = L(\boldsymbol{\theta} \mid \mathbf{X}, \mathbf{Y}) = f(\mathbf{X}, \mathbf{Y} \mid \boldsymbol{\theta})$ and the observed data likelihood is

$$L(\boldsymbol{\theta} \mid \mathbf{Y}) = \int f(\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\theta}) f(\mathbf{X} \mid \boldsymbol{\theta}) \partial \mathbf{X}.$$

The EM algorithm first finds the expected value of the complete-data log-likelihood with respect to the unknown data, $\mathbf{X}$, given the observed data, $\mathbf{Y}$, and the current parameter estimates in the E-step:

$$Q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)}\right) = E\left[\log f(\mathbf{X}, \mathbf{Y} \mid \boldsymbol{\theta}) \mid \mathbf{Y}, \boldsymbol{\theta}^{(m-1)}\right]$$

where $\boldsymbol{\theta}^{(m-1)}$ are the current parameter estimates used to evaluate the expectation and $\boldsymbol{\theta}$ are the new parameters that are optimized to increase Q. Secondly, in the M-step, $\boldsymbol{\theta}^{(m)}$ are chosen to maximize the expectation such that: $Q\left(\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(m-1)}\right) \geq Q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m-1)}\right)$. The algorithm is iterated until the difference, $L\left(\boldsymbol{\theta}^{(m)}\right) - L\left(\boldsymbol{\theta}^{(m-1)}\right)$ is sufficiently small.

McLachlan and Krishnan (1997) list some of the nice properties of the EM compared to other iterative algorithms such as Newton-Raphson and Fisher's scoring

methods. Importantly, the EM is numerically stable, with each iteration increasing the likelihood; and in general, the global convergence is reliable, that is, starting from an arbitrary point in the parameter space, convergence almost always occurs to a local maximizer. Wu (1983) presents a detailed study of the convergence properties of the EM algorithm.

One of the criticisms commonly associated with the EM algorithm is that it does not have a built-in procedure to automatically produce standard errors of the parameter estimates. However, this issue has been addressed by different authors. Louis (1982) developed a method for estimating the standard errors based on the observed information matrix using score equations. Efron developed methods for constructing approximate confidence intervals using the bootstrap approach (1986, 1987). The bootstrap is a general technique for estimating unknown quantities associated with statistical models. It is typically used to estimate quantities associated with the sampling distribution of estimators and test statistics [Boos 2003]. Details of the bootstrap procedure for percentile and bias-corrected confidence interval estimation were provided in Chapter 4.

EM Algorithm Application

It is assumed that an individual is a responder with probability p in which case $Z_i=1$ or a non-responder with probability (1-p) in which case $Z_i=0$. For responders, we assume $\mathbf{v}_i = \left(K_i^*, Km_i^*\right)^T \sim MVN(\boldsymbol{\mu}_1, \Sigma_1)$ and, for non-responders, we assume $\mathbf{v}_i = \left(K_i^*, Km_i^*\right)^T \sim MVN(\boldsymbol{\mu}_2, \Sigma_2),$ with $\boldsymbol{\mu}_r = \left(\kappa_r, \kappa m_r\right)^T$ and

$$\Sigma_r = \begin{pmatrix} \sigma^2_{K^*_p} & \sigma_{K^*_p,Km^*_p} \\ \sigma_{K^*_p,Km^*_p} & \sigma^2_{Km^*_p} \end{pmatrix}, \quad \text{for } r=1, 2.$$ The likelihood of this mixture distribution,

conditional on the responder status, can be specified as

$$\prod_{i=1}^{m} \left\{ p \int f(\mathbf{v}_i | \mu_1, \Sigma_1) \right\}^{Z_i} \left\{ (1-p) \int f(\mathbf{v}_i | \mu_2, \Sigma_2) \right\}^{1-Z_i} \tag{5.3}$$

where $f(\mathbf{v}_i | \mu_1, \Sigma_1)$ and $f(\mathbf{v}_i | \mu_2, \Sigma_2)$ are the multivariate normal densities specified

previously for responders and non-responders, respectively. In general we may assume,

p, the probability of being in the responder group, can be estimated using a logistic

regression model where $p = \dfrac{e^{\alpha+\beta a}}{1+e^{\alpha+\beta a}}$ with a representing a covariate(s) of interest. In

order to accommodate the unknown responder status within the estimation procedure, an

EM algorithm was employed.

Both the observed DE data and the simulated mixture distribution data generated for

Chapter 3 were processed through the EM algorithm to estimate parameters. The EM

algorithm was used to find the maximum likelihood estimates of the parameters

p, μ_1, Σ_1, μ_2, Σ_2, and Z_i through iteration of the expectation and maximization steps.

An outline of the steps comprising the EM algorithm is provided below:

a. Provide initial values for the unknown parameters, p, μ_1, Σ_1, μ_2, Σ_2

b. Perform the expectation step to estimate the expected value of Z_i

$$e_{i1} = E\left(Z_i \mid y_i, p^{(m)}, \mu_1^{(m)}, \Sigma_1^{(m)}, \mu_2^{(m)}, \Sigma_2^{(m)}\right) =$$
$$\left. p^{(m)} \phi\left(y_i; \mu_1^{(m)}, \Sigma_1^{(m)}\right) \middle/ p^{(m)} \phi\left(y_i; \mu_1^{(m)}, \Sigma_1^{(m)}\right) + \left(1-p^{(m)}\right) \phi\left(y_i; \mu_2^{(m)}, \Sigma_2^{(m)}\right) \right.$$

$$e_{i2} = 1 - e_{i1}$$

Define $\mathbf{e_1}=(e_{11},...,e_{m1})^T$ and $\mathbf{e_2}=(e_{12}, ..., e_{m2})^T$, $\mathbf{y}=(\mathbf{y_1},...,\mathbf{y_m})^T$ for $\mathbf{y_i} = \left(K_i^*, Km_i^*\right)$ and

$\mathbf{I_m}$ is an mxm identity matrix.

c. Perform the maximization steps based on the log likelihood:

$$\log L = \sum_{i=1}^{m} Z_i \log p + Z_i \log\left\{\phi\left(y_i; \mu_1^{(m)}, \Sigma_1^{(m)}\right)\right\} +$$
$$\left(1 - Z_i\right)\log\left(1 - p\right) + \left(1 - Z_i\right)\log\left\{\phi\left(y_i; \mu_2^{(m)}, \Sigma_2^{(m)}\right)\right\}$$

i) $\quad p^{(m+1)} = \sum_{i=1}^{m} \dfrac{e_{i1}}{m}$

ii) $\quad \mu_1^{(m+1)} = \dfrac{\mathbf{e_1^T} \times \mathbf{y}}{\sum\limits_{i=1}^{m} e_{i1}}$

iii) $\quad \mu_2^{(m+1)} = \dfrac{\mathbf{e_2^T} \times \mathbf{y}}{\sum\limits_{i=1}^{m} e_{i2}}$

iv) $\quad \Sigma_1^{(m+1)} = \left(\mathbf{y} - \mu_1^{(m+1)}\right)^T \mathbf{e_1} \mathbf{I}_m \left(\mathbf{y} - \mu_1^{(m+1)}\right)$

v) $\quad \Sigma_2^{(m+1)} = \left(\mathbf{y} - \mu_2^{(m+1)}\right)^T \mathbf{e_2} \mathbf{I}_m \left(\mathbf{y} - \mu_2^{(m+1)}\right)$

d. Iterate until convergence is achieved.

Following convergence, the times of maximum excretion for the responder and non-responder groups, *T1* and *T2* were calculated using the μ_1 and μ_2 estimates provided by the EM algorithm. Confidence intervals for *T1* and *T2*, using the first order delta method cited in Chapter 4, were calculated based on estimated Σ_1 and Σ_2. The bootstrap method with b=250 bootstrap samples was used to provide 90% percentile confidence intervals for the p, μ_1, Σ_1, μ_2, and Σ_2 parameters estimated by the EM algorithm.

Results

Observed Data

Although the observed data appeared as a mixture of multiple distributions, review of the individual subject K_i^* and Km_i^* values plotted against the estimated Tm_i (Figure 5.1) showed that although Tm_i values ranged from approximately 1 to 23 hours, there was little variation between K_i^* and Km_i^* associated with these different Tm_i's. Attempts to identify a mixture distribution within the EM algorithm were not successful, except the separation of the single highest Tm_i from the remainder of the Tm_i estimates.
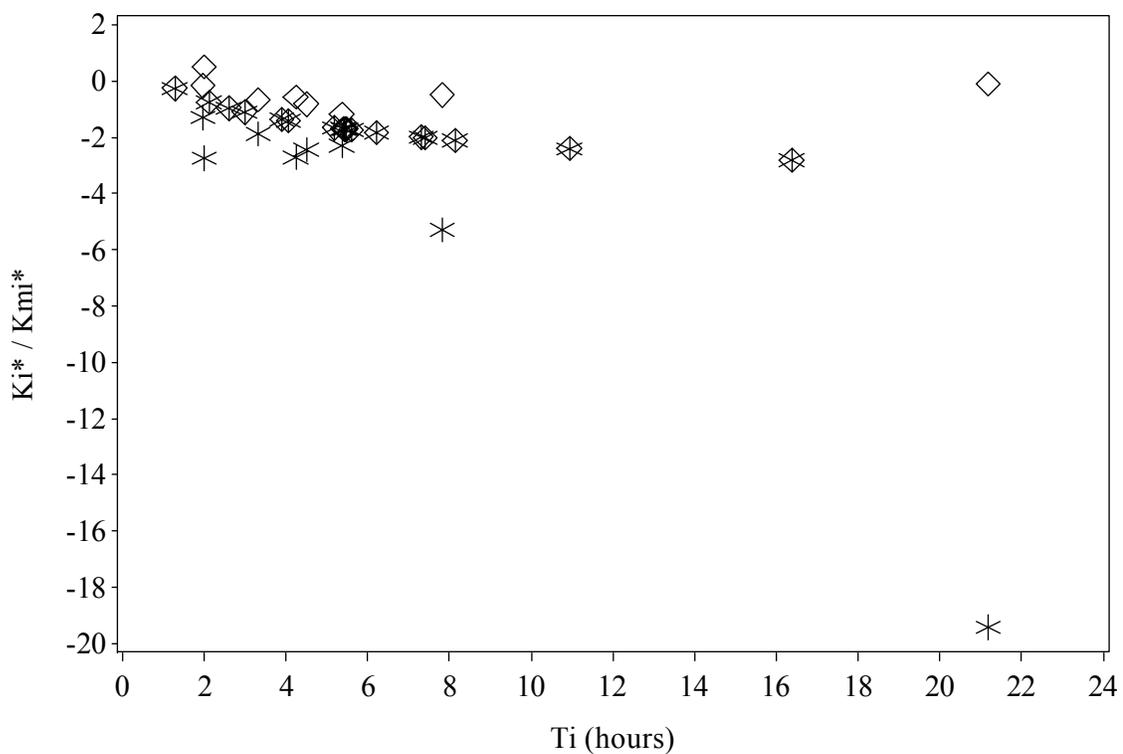
Figure 5.1: Scatter Plot of estimated K_i^* (*) and Km_i^* (◊) by Tm_i for observed 1-AP urine excretion data in 25 subjects.

Simulated Data

Table 5.1 shows the predicted values and bootstrap confidence intervals (CI) for

the "responder" and "non-responder" groups in the simulated mixture data set. The

estimated p was 0.50 (90% CI: 0.32 – 0.65), similar to the "true' p of 0.50. *T1* and *T2*

differed somewhat from the "true" values of 1.51 and 3.74 hours. Ninety-five per cent

CIs generated via the first order delta method were 0.50 to 3.78 hours for *T1*

("responders") and 0.87 to 5.62 hours for *T2* ("non-responders").

Table 5.1: Results of the EM Algorithm for the Simulated Data

	Group 1				Group 2		
	"True" Value	EM Estimate	Bootstrap 95% CI		"True" Value	EM Estimate	Bootstrap 95% CI
Tmax	1.51	2.26	1.52 - 2.43		3.74	3.09	2.96 – 4.43
K*	-0.5	-1.08	-1.22 - -0.50		-2.1	-1.63	-2.20 - -1.52
Km*	-0.33	-0.57	-0.60 - -0.32		-0.7	-0.69	-0.93 - -0.66
$\sigma^2_{K*,K*}$	0.03	7.55	0.05 – 7.91		0.03	7.52	0.07 - 7.93
$\sigma^2_{Km*,Km*}$	0.12	1.60	0.51 – 1.90		0.12	1.57	0.46 – 1.99
$\sigma_{K*,Km*}$	0.0024	1.92	0.03 – 2.17		0.0024	1.92	-0.00 – 2.10

Figure 5.2 shows the K_i* and Km_i* versus Tm_i values for the simulated data. There was a

more distinct difference in the K_i* versus Tm_i values for the 2 groups in the simulated

mixture distribution compared to the observed data displayed in Figure 5.1.
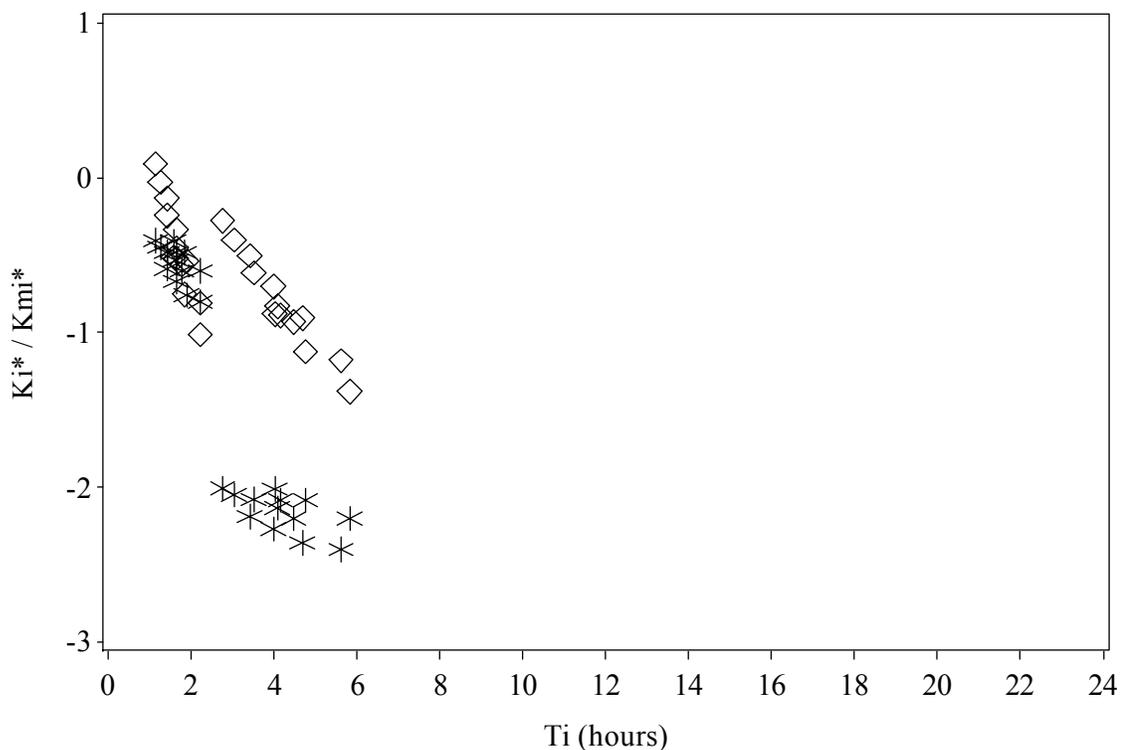
Figure 5.2: Scatter Plot of estimated K_i^* (*) and Km_i^* (◊) by Tm_i for Simulated 1-AP urine excretion data.

Discussion

In this chapter we have explored an application of the EM algorithm in an attempt to further characterize the time of maximum 1-AP excretion following a controlled DE exposure. Previous work indicated that a mixture of more than one response profile may be present in the experimental subject's data, which may partially explain the differences in responses. While a mixture may be present, in that some subjects may exhibit Tm_i within 24-hours after exposure and others later than 24-hours following exposure, insufficient data on the latter group of subjects prevented further exploration of this hypothesis. Exploration of a mixture distribution present in the data where Tm_i occurred earlier than 24-hours following exposure did not indicate that a mixture of two

distributions was present in the 1-AP data, but rather that the distribution of Tm_i among the subjects was a continuum of values spread over a range varying from 1 to 16+ hours, with a single subject's Tm_i estimated at approximately 21 hours.

The observed DE experimental data was challenging to model due not only to the wide variety of subject responses, but also due to the limited amount of information available for each subject. In PK applications where individual responses are more conforming to a general trend, sparse data may be combined for nonlinear modeling using population-average approaches. This was not possible in the observed 1-AP data, resulting in the ability to model individual subject profiles only.

The use of a simulated mixture distribution data set demonstrated that an EM algorithm based on a multivariate normal distribution for the PK parameter estimates was able to distinguish two distributions, however, not as well as expected. This approach may be a viable strategy for future exploration of mixture distributions in PK applications.

References

Bamford, H.; Baker, J. Nitro-polycyclic aromatic hydrocarbon concentrations and sources in urban and suburban atmospheres of the Mid-Atlantic region. *Atmos. Environ. 2003. 37(15),* 2077-2091.

Beland, F. A.; Kadlubar, F. F. Metabolic Activation and DNA Adducts of Aromatic Amines and Nitroaromatic Hydrocarbons. In *Carcinogenesis and Mutagenesis I* ; Cooper, C.S.; Grover, P.L., Eds.; Springer-Verlag: New York, 1990, 267-325.

Bilmes, J. A. (1998). A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models (http://citeseer.ist.psu.edu/bilmes98gentle.html)

Bond, J. A., Sun, J. D., Medinsky, M. A., Jones, R. K. , Yeh, H. C. (1986). Deposition, Metabolism, and Excretion of 1-[^{14}C]Nitropyrene and 1-[^{14}C]Nitropyrene Coated on Diesel Exhaust Particles as Influenced by Exposure Concentration. *Toxicology and Applied Pharmacology.* 85, 102-117.

Boos, D. D. & Stefanski, L. A. (2006). *Modern Statistical Inference: Theory and Methods.* New York: Springer.

Davidian, M. & Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data.* Boca Raton: Chapman & Hall.

Dempster, A., Laird, N., Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B,* 39 (1) 1-38.

Dutcher, J. S., Sun, J. D., Bechtold, W. E., Unkefer, C. J. (1985). Excretion and Metabolism of 1-Nitropyrene in Rats after Oral or Intraperitoneal Administration. *Toxicological Sciences*, 5 (2), 287-296.

Efron, B. & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science,* 1 (1) 54-77.

Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association,* 82 (397) 171-185.

Howard, P. C., Consolo, M. C., Dooley, K. L., Beland, F. A. (1995). Metabolism of 1-Nitropyrene in Mice: Transport Across the Placenta and Mammary Tissues. *Chemico-Biological Interactions,* 95, 309-325.

IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. http://monographs.iarc.fr/ENG/Monographs/vol46/volume46.pdf (accessed July

30, 2008). Volume 46 Diesel and Gasoline Engine Exhausts and Some Nitroarenes.

IPCSINTOXDatabank. http://www.intox.org/databank/documents/chemical/nitrooxy/ehc229.htm. (accessed March 28, 2009).

Laumbach, R. (2009). Quantification of 1-aminopyrene in human urine after a controlled exposure to diesel exhaust. *Journal of Environmental Monitoring*, (11) 153-159.

Louis, Thomas A. (1982). Finding the Observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society*, (44) 226-233.

McLachlan, G. J. & Krishnan, T. (1997). *The EM Algorithm and Extensions.* New York: John Wiley & Sons, Inc.

SAS Institute Inc. *SAS OnlineDoc®,* Version 9.1. SAS Institute Inc.: Cary, NC, 2004.

Sun, J. D., Wolff, R. K. , Aberman, H. M. , McClellan, R. O. (1983). Inhalation of 1-nitropyrene Associated with Ultrafine Insoluble Particles or as a Pure Aerosol: A Comparison fo Deposition and Biological Fate. *Toxicology Applied Pharmacology,* 69 (2) 185-198.

Tokiwa, H.; Ohnishi, Y. Mutagenicity and carcinogenicity of nitroarenes and their sources in the environment. *Crit Rev Toxicol.* 1986*. 17(1),* 23-60.

Wu, C. F. J. (1983). On the Convergence Properties of the EM Algorithm. *Annals of Statistics,* 11, 95-103.

Chapter 6

Discussion and Recommendations

This project was begun with the goal of identifying and characterizing the expected time of maximum urine excretion (Tmax) of a diesel exhaust biomarker, 1-aminopyrene. Reaching this goal would allow for enhancement of future experimental designs and potentially simplified, targeted sample collections. Challenges to this goal arose due to the variety of subject response profiles. To meet this challenge, the estimation of the time of maximum urine excretion of diesel exhaust biomarker 1-aminopyrene included: (1) identification of the best-fitting pharmacokinetic model; (2) evaluation of the distribution of the urine concentrations over time and the impact of factors such as peak excretion time, pharmacokinetic (PK) parameter variability, and sampling scheme; (3) comparison of PK parameter estimation methods and confidence interval approaches; and (4) development of an EM algorithm for characterization of a mixture of subject response profiles.

The first task of this dissertation was to identify the best-fitting PK model for the urine excretion data. While there are advantages to using urine data over plasma data, such as less invasive sample collection and the ability to collect and store samples away from the clinic site, disadvantages also exist, such as difficulty in obtaining specifically timed or frequent samples as well as complete voids. Other differences exist, for example, plasma concentrations represent the amount of a drug or chemical in the plasma at the time of sample collection, while urine concentrations represent the cumulative amount of a drug or chemical collected since the previous void. Based on these

differences, PK models based on the rate of excretion as well as the cumulative excretion were compared to evaluate fit. Results of an analysis that considered the distribution, location, and variance patterns of the residuals, showed that the cumulative distribution model provided the best data fit.

The bulk of the work in this dissertation concerned the evaluation of non-compartmental and nonlinear modeling methods along with the comparison of several standard confidence interval techniques, (non-parametric, standard asymptotic, and bootstrap) along with a proposed confidence interval based on the delta method. In the author's review of the literature, no prior work was found comparing the non-compartmental, global two-stage, and numerical approximation (Laplace method) modeling approaches using simulated data. This exercise demonstrated the superiority of numerical approximation using the Laplace method in terms of accuracy and precision of the PK parameter estimates as well as the ease of use (relative to the global two-stage method). As with the PK modeling comparison, in the author's review, no comprehensive comparison of confidence interval methods for single sample Tmax were identified in the literature. Comparison of the confidence interval methods showed that when only individual subject-level estimates of Tmax were obtainable, the non-parametric confidence interval method was optimal because it was most likely to provide coverage at the nominal $100(1-\alpha)\%$ level, where α is the type I error rate. If an overall, population-average Tmax was estimable from the data, the first-order delta method confidence interval performed the best. This method provided coverages close to the nominal level.

These conclusions regarding the nonlinear modeling and confidence interval approaches were determined based on simulated data created to mimic the observed 1-AP values. Modeling of the observed data proved difficult due to the wide variety of subject response profiles. In addition a loss of data occurred due to several issues. For example, of the 55 original subjects, 15 were excluded from any primary analyses due to missing or invalid samples. Of the remaining 40 subjects, approximately 30% had late peaks occurring near the end of the planned collection period, making assessment of Tmax impossible. This loss of data resulted in difficulty conclusively attaining our original research goal of characterizing the time of maximum 1-AP excretion for simplification of sample timing in future studies. Nevertheless, we were able to demonstrate the statistical methods used in this dissertation on the observed 1-AP data, and made important methodological comparisons using simulated data.

With the variety of observed responses and through the investigation of various factors such as PK parameter value, variability and timing of sample collection, it was hypothesized that the observed 1-AP data may represent a mixture of more than one distribution. This is biologically plausible, as previous work has cited polymorphisms which affect the metabolism of 1-nitropyrene, the parent compound of 1-AP, present in diesel exhaust [IPCSINTOXDatabank]. Exploration of this mixture distribution was difficult due to the inability to model the subjects with potentially late excretion peaks (e.g. later than the 24-hour collection period). Demonstration of the EM algorithm using the multivariate distribution of the Tmax PK parameters for absorption (metabolism) and excretion rates was done with simulated data.

Conclusions and Recommendations

While challenges existed with the 1-AP urine excretion data which could not be overcome with statistical analysis, such as lack of sample collection after 24 hours for potential late-peak profiles, and invalid creatinine levels leading to "unusable" samples resulting in missing values, 25 subjects did provide usable data from which information for characterizing the time of maximum 1-AP excretion could be derived. Specifically, using the recommended non-parametric confidence interval technique for these 25 subjects, we found that collection of voids occurring within an approximate 3.9 to 6.2 hour window following exposure should, with 95% confidence, provide assays that represent the median peak urine excretion of the DE biomarker 1-AP in subjects who have maximum excretions within 24 hours following exposure.

In addition, important insights to the modeling, parameter estimation technique, and confidence interval approach were gained. In particular, recommendations for nonlinear modeling using the Laplace method and a cumulative excretion model for urine excretion data were identified. In addition, when individual subject-level PK parameters were estimated and Tmax calculated from these, non-parametric confidence intervals were noted to have provided coverages closest to the nominal level, given a sufficient number of subjects were present in the sample. Further, when population-average modeling techniques were used, for example, in the case of sparse sampling, the first order delta method was noted to have provided coverages closest to the nominal level. This was useful information for future studies, in general, as the collection of urine samples is an important tool for public health studies which evaluate exposures to toxic substances through biomarker analysis.

Recommendations for Future Work

Future experimental work with DE biomarkers should include a study, similar to the one cited in this dissertation utilizing a controlled DE exposure, but with an extended urine collection period, beyond 24 hours so that subjects who may be excreting these DE biomarker substances later than 1 day following exposure can be characterized more fully. Animal studies [Bond, et al 1986] showed that following exposure to radio-labeled DE, half-times for elimination in urine and feces ranged from 15 to 20 hours. Collecting samples for 2 to 3 days after exposure would be expected to cover approximately 3 to 5 half-times for elimination and thus provide a reasonable expectation of more fully characterizing the time course of elimination in humans.

Future statistical work on this topic should include an exploration of the impact of measurement error from the individual subject-level Tmax estimates on the confidence intervals for percentiles that are not near the center of the distribution. For example, in the Chapter 4 Discussion section, we sought to answer the question "How long after an exposure should samples be collected in order to assure 90% probability of collecting at least 80% of the individual subject-level Tmax values?" and provided an estimate of at least between 5.5 to 16.4 hours; however, methods have been published which may help to mitigate the effect of measurement error on the confidence interval and should be explored further [Schechtman and Spiegelman 2007].

Public Health Recommendations

Health risks associated with from DE exposure have been well-documented in the literature and the media. Direct (occupational) exposure to diesel exhaust is associated with increased risk of lung cancer [National Institute for Occupational Safety and Health] and respiratory diseases and symptoms [Pandya 2002]. In addition, nitrogen oxides from diesel emissions add to the formation of ground level ozone that can lead to irritation of the respiratory system, resulting in coughing, choking, and reduced lung capacity. In addition, urban ozone pollution has been associated with increased hospitalizations for respiratory illnesses such as asthma.

DE has been classified a potential human carcinogen by the U.S. Environmental Protection Agency (EPA) and the International Agency for Research on Cancer. Studies of humans routinely exposed to diesel fumes point to a greater risk of lung cancer. For example, occupational health studies of workers exposed to high levels of diesel exhaust over many years consistently demonstrate a increased the risk of lung cancer or mortality of 20 to 50 percent. [Health Assessment Document for Diesel Engine Exhaust].

Sources of DE are ubiquitous in the United States; they include not only many of the buses and trucks in our cities and on our highways but also marine engines as well as farm and construction equipment [http://www.edf.org/ documents]. Thus, DE exposure is a concern for both urban and non-urban dwelling people. While steps are being taken to alleviate DE exposure through the development of alternate fuels, the elimination of DE exposure will take time. Methods for the accurate, convenient, and efficient measuring of DE exposure are needed to aid in biomonitoring of this public health risk.

Based on the results of DE exposure study cited in this dissertation and the work of this dissertation, the following recommendations can be made:

1. Urine biomonitoring is a viable option for detection of DE exposure;

2. Samples for 1-AP should include voids occurring up until approximately 5.5 to 16.4 hours after DE exposure to assure with 90% probability that at least 80% of the individuals sampled will have obtained peak 1-AP excretion; and,

3. Today there are state and/or federal legislations aimed at retrofitting existing diesel technologies to reduce diesel exhaust emissions. Simultaneously new diesel engine technology has been introduced with substantially reduced emissions of DE particles. In order to evaluate the effectiveness of these legislative actions and the effectiveness of the technologies, it is important to accurately measure changes in DE exposure brought by these interventions. Using urinary 1-AP in a biomonitoring program is recommended, preferably after the remaining issues of this biomarker, as discussed in this dissertation, have been clarified or resolved.

.

References

Bond, J. A., Sun, J. D., Medinsky, M. A., Jones, R. K. , Yeh, H. C. (1986). Deposition, Metabolism, and Excretion of 1-[^{14}C]Nitropyrene and 1-[^{14}C]Nitropyrene Coated on Diesel Exhaust Particles as Influenced by Exposure Concentration. *Toxicology and Applied Pharmacology.* 85, 102-117.

Closing the Diesel Divide, http://www.edf.org/documents/2738_DieselDivide.pdf (accessed October 23, 2009).

Health Assessment Document for Diesel Engine Exhaust. National Center for Environmental Assessment, Office of Research and Development, US EPA. Washington D.C. May 2002. page 9-11. EPA/600/8-90/057F

IPCSINTOXDatabank. http://www.intox.org/databank/documents/chemical/nitrooxy/ehc229.htm. (accessed March 28, 2009).

National Institute for Occupational Safety and Health (NIOSH). Carcinogenic Effects of Exposure to Diesel Exhaust. Current Intelligence Bulletin 50. 1988. http://www.cdc.gov/niosh/88116_50.html. (accessed November 11, 2009).

Pandya,R. J., Solomon, G., Kinner, A., Balmes, J. R. (2002). Dielse Exhaust and Asthma: Hypotheses and Molecular Mechanisms of Action. *Environmental Helath Perspectives*. 110(1) 103-112.

Schechtman, E., Spiegelman, C. (2007). Mitigating the effect of measurement errors in quantile estimation. *Statistics & Probability Letters*, 77, 514-524.

BIBLIOGRAPHY

Bamford, H.; Baker, J. Nitro-polycyclic aromatic hydrocarbon concentrations and sources in urban and suburban atmospheres of the Mid-Atlantic region. *Atmos. Environ. 2003. 37(15),* 2077-2091.

Barr, D.B.; Wang, R.Y.; Needham,L.L. Biologic Monitoring of Exposure to Environmental Chemicals throughout the Life Stages: Requirements and Issues for consideration for the National Children's Study. *Environ. Health Perspect.* 2005. *113(8),* 1083-1091.

Beland, F. A.; Kadlubar, F. F. Metabolic Activation and DNA Adducts of Aromatic Amines and Nitroaromatic Hydrocarbons. In *Carcinogenesis and Mutagenesis I* ; Cooper, C.S.; Grover, P.L., Eds.; Springer-Verlag: New York, 1990, 267-325.

Bilmes, J. A. (1998). A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models (http://citeseer.ist.psu.edu/bilmes98gentle.html)

Bond, J. A., Sun, J. D., Medinsky, M. A., Jones, R. K. , Yeh, H. C. (1986). Deposition, Metabolism, and Excretion of 1-[^{14}C]Nitropyrene and 1-[^{14}C]Nitropyrene Coated on Diesel Exhaust Particles as Influenced by Exposure Concentration. *Toxicology and Applied Pharmacology.* 85, 102-117.

Boos, D. D. & Stefanski, L. A. (2006). *Modern Statistical Inference: Theory and Methods.* New York: Springer.

Cardiello, P. Kroon, E. oeelmans, R., Vorarien, W., Ubolyam, S., Phanuphak, P., Lange, J., Cooper, D., Ruxrungthan, K. (2000). Steady-state pharmacokinetics (PK) of saquinavir (SQV) with and without two doses of itraconazole in HIV-1+ Thai patients:HIV NAT001.2 substudy. Int Conf AIDS, abstract no. WeOrB545.

Charles, B.G., Miller, A.K., Nasveld, P.E., Reid, M.G., Harris, I.E., Edstein, M.D. (2007). Populations Pharmacokinetics of Tafenoquine during Malaria Prophylaxis in Healthy Subjects. *Antimicrobial Agents and Chemotherapy*, 2709-2715.

Chow, S-C & Liu, J-P. (1992). *Design and Analysis of Bioavailability and Bioequivalence Studies*. New York: Marcel Dekker, Inc.

Closing the Diesel Divide, http://www.edf.org/documents/2738_DieselDivide.pdf (accessed October 23, 2009).

Cooch, E.G. & White, G.C. [online]. 2009. The Delta method http://www.phidot.org/software/mark/docs/book/pdf/app_2.pdf (last accessed June 2009).

Cornell, R. G. (1991). Nonparametric Tests of Dispersion for the Two-Period Crossover Design. *Commun. Statist. Theory Meth*. 20(3), 1099-1106.

Davidian, M. & Giltinan, D. M. (1993). Some Simple Methods for Estimating Intra-individual Variability in Nonlinear Mixed Effects Models. *Biometrics,* 49, 59-73.

Davidian, M. & Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data.* Boca Raton: Chapman & Hall.

Davidian, M. & Giltinan, D. M. (2003). Nonlinear Models for Repeated Measurement Data: An Overview and Update. *Journal of Agricultural, Biological & Environmental Statistics, 8* (4) 387-419.

Dempster, A., Laird, N., Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B,* 39 (1) 1-38.

Dutcher, J. S., Sun, J. D., Bechtold, W. E., Unkefer, C. J. (1985). Excretion and Metabolism of 1-Nitropyrene in Rats after Oral or Intraperitoneal Administration. *Toxicological Sciences*, 5 (2), 287-296.

Efron, B. & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science,* 1 (1) 54-77.

Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association,* 82 (397) 171-185.

Goodman, L. A. (1962). The Variance of the Product of K Random Variables. *Journal of the American Statistical Association*, 57(297) 54-60.

Hall, P., (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hauschke, D. Steinijans, V.W., Diletti, E. (1990). A distribution-free procedure for the statistical analysis of bioequivalence studies. *International Journal of Clinical Pharmacology, Therapy and Toxicology*, (28) 72-78.

Health Assessment Document for Diesel Engine Exhaust. National Center for Environmental Assessment, Office of Research and Development, US EPA. Washington D.C. May 2002. page 9-11. EPA/600/8-90/057F

Hogg, R. V. & Craig, A. T. (1978) *Introduction to Mathematical Statistics*. Macmillan Publishing Co., Inc.: New York.

Howard, P. C., Consolo, M. C., Dooley, K. L., Beland, F. A. (1995). Metabolism of 1-Nitropyrene in Mice: Transport Across the Placenta and Mammary Tissues. *Chemico-Biological Interactions,* 95, 309-325.

IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. http://monographs.iarc.fr/ENG/Monographs/vol46/volume46.pdf (accessed July 30, 2008). Volume 46 Diesel and Gasoline Engine Exhausts and Some Nitroarenes.

IPCSINTOXDatabank. http://www.intox.org/databank/documents/chemical/nitrooxy/ehc229.htm. (accessed March 28, 2009).

Islinger, F., Dehghanyar, P., Sauermann, R., Burger, C., Kloft, C., muller, M., Joukhadar, C. (2006). The effect of food on plasma and tissue concentrations of linezolid after multiple doses. *International Journal of Antimicrobial Agents*, 27, 108-112.

Laumbach, R. (2009). Quantification of 1-aminopyrene in human urine after a controlled exposure to diesel exhaust. *Journal of Environmental Monitoring*, (11) 153-159.

Lindstrom, M. & Bates, D. (1990). Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics,* (46) 673-687.

Louis, Thomas A. (1982). Finding the Observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society*, (44) 226-233.

McLachlan, G. J. & Krishnan, T. (1997). *The EM Algorithm and Extensions.* New York: John Wiley & Sons, Inc.

Mortimer, K.J., Tattersfield, A.E., Tang, Y., Wu, K., Hochhaus, S. , Harrison, T.W. (2007). Plasma concentrations of Fluticasone propionate and budesonide following inhalation: effect of induced bronchoconstriction, *British Journal of Clinical Pharmacology*, 64(4), 439-444.

National Institute for Occupational Safety and Health (NIOSH). Carcinogenic Effects of Exposure to Diesel Exhaust. Current Intelligence Bulletin 50. 1988. http://www.cdc.gov/niosh/88116_50.html. (accessed November 11, 2009).

Pandya,R. J., Solomon, G., Kinner, A., Balmes, J. R. (2002). Dielse Exhaust and Asthma: Hypotheses and Molecular Mechanisms of Action. *Environmental Helath Perspectives*. 110(1) 103-112.

Pinheiro, J. C. & Bates, D. M., (1995). Approximations to the Log-Likelihood function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics,* 4 (1) 12-35.

Pinheiro, J. C. & Bates, D. M., (2004). *Mixed Effects Models in S and S-Plus.* New York: Springer Science+Business Media, LLC.

Saito, K.; Kamataki, T.; Kato, R. Participation of Cytochrome P-450 in Reductive Metabolism of 1-Nitropyrene by Rat Liver Microsomes. *Cancer Res*. 1984, 44, 3169-3173.

SAS Institute Inc. *SAS OnlineDoc®,* Version 9.1. SAS Institute Inc.: Cary, NC, 2004.

Schechtman, E., Spiegelman, C. (2007). Mitigating the effect of measurement errors in quantile estimation. *Statistics & Probability Letters*, 77, 514-524.

Seidel, A.; Dahmann, D.; Krekeler, H.; Jacob, J. Biomonitoring of polycyclic aromatic compounds in the urine of mining workers occupationally exposed to diesel exhaust. *Int. J. Hyg. Environ. Health.* 2002. *204,* 333-338.

Shao, J. & Tu, D., (1995). *The Jackknife and Bootstrap.* New York: Springer-Verlag.

Shargel, L. & Yu, A. B. C. (1993). *Applied Biopharmaceutics and Pharmacokinetics.* Stamford: Appleton & Lang.

Sheiner, L. B. & Beal, S. L. (1980). Evaluation of Methods for Estimating Population Pharmacokinetic Parameters. I. Michaelis-Menten Model: Routine Clinical Pharmacokinetic Data. *Journal of Pharmacokinetics and Biopharmaceutics,* 8, 553-571.

Smith, P.F., Forrest, A., Ballow, C.H., Adams, J.M., Shiveley, L.R. (2002). Single dose pharmacokinetics of SPD-756 in healthy adult volunteers, Int. Conf. AIDS, abstract no. WePeB6050.

S-Plus 6.0 Professional Release 2 for Windows

Sun, J. D., Wolff, R. K. , Aberman, H. M. , McClellan, R. O. (1983). Inhalation of 1-nitropyrene Associated with Ultrafine Insoluble Particles or as a Pure Aerosol: A Comparison fo Deposition and Biological Fate. *Toxicology Applied Pharmacology,* 69 (2) 185-198.

Tokiwa, H.; Ohnishi, Y. (1986). Mutagenicity and carcinogenicity of nitroarenes and their sources in the environment. *Crit Rev Toxicol.. 17(1),* 23-60.

Triantafyllopoulos K., (2003). On the central moments of the multidimensional Gaussian distribution. The Mathematical Scientist, 18 125-128.

Vonesh, E. F. (1996). A Note on the Use of Laplace's Approximation for Nonlinear Mixed-Effects Models. *Biometrika,* 83, 447-452.

Vonesh, E. & Carter, R. (1992). Mixed-Effects Nonlinear Regression for Unbalanced Repeated Measures. *Biometrics,* (48) 1-17.

Willavize, S.A. & Morgenthien, E. A. (2008). Nonparametric confidence intervals for Tmax in sequence-stratified crossover studies. *Pharmaceutical Statistics*, (7), 9-19.

Wolfinger, R. (1993). Laplace's Approximation for Nonlinear Mixed Models. *Biometrika*, 80, 791-795.

Wu, C. F. J. (1983). On the Convergence Properties of the EM Algorithm. *Annals of Statistics,* 11, 95-103.

CURRICULUM VITA

EDUCATION: *Master of Science, Biostatistics*, 1983
               COLUMBIA UNIVERSITY, School of Public Health,
               New York, New York
               Bachelor of Science, Biology, Cum Laude, 1982
               FAIRFIELD UNIVERSITY, Fairfield, Connecticut

WORK EXPERIENCE:

SCHERING PLOUGH RESEARCH INSTITUTE   1999-Present
   Manager, Statistics
   • Supervise statistical staff in support of drug development including creating the
     Statistical Analysis Plan (SAP), performing statistical analyses and reporting
     results to upper management, clinical team, and outside experts
   • Determine the statistical methodology for clinical study protocols; collaborate
     with clinical project team on protocol design and development as well as overall
     product development plans
   • Represent Statistics department in cross-functional therapeutic and study team
     meetings
   • Review and approve clinical study reports
   • Create, manage, and release random codes, as necessary
   • Represent Statistics department on Due Diligence visits and report findings to
     upper management
   • Member of the organizational committee and session leader for CNS in
     FDA/Industry Statistics Workshop 2009
   Consultant Statistician
   • Recommend study design and analysis methodology and perform planned and ad-
     hoc statistical analyses supporting non-clinical drug development using SAS (e.g.
     Proc Mixed, Proc GLM, Proc Freq, Proc Capability);
   • Write reports of statistical findings including graphical presentation of data using
     SAS Graph; Collaborate with medical/scientific personnel to incorporate these
     statistical findings in scientific reports;
   • Assist in development of product specifications using tolerance intervals,
     regression analyses and data prediction models;
   • Participate with medical/scientific personnel in the review of protocols and
     writing of statistical sections of protocols, including study design and sample size
     considerations;
   • Represent statistics group in scientific drug development teams including joint
     ventures with other pharmaceutical companies and vendors;
   • Review literature for statistical methodologies and be involved in the
     development of such methodology.

VARIOUS PHARMACEUTICAL COMPANIES   1995-1999
   Data Management Consultant / SAS Programmer

- Analyzed data and prepared summary data displays for various projects including: integrated safety summary; supplemental integrated summary of efficacy; supplemental NDA, and NDAs- all programmed in SAS.
- Performed ad hoc safety data analysis for marketing study
- Coordinated data mapping of over 75 completed clinical trials into a single data base using a common data dictionary.

BOEHRINGER INGELHEIM, Ridgefield, Connecticut   1985 - 1994
SWF Project Data Manager (1989-1994)

- Worked with multiple clinical project teams in all phases of product development including protocol review, monitoring trial conduct, final reports and NDA preparation.
- Managed databases for Phase II and III clinical trials.
- Trained new SAS programmers and clinical data analysts.
- Created data displays and generated reports using SAS, including periodic data reviews and annual safety updates.

Supervisor, Clinical Trials Information (1985-1989)
- Supervised and trained clinical trial analysts in the review, management and reporting of clinical trial data.
- Monitored consistency and accuracy of assigned drug databases.
- Developed and reviewed clinical trials and screening manuals for clinical trial management.
- Served as a member of clinical monitoring team overseeing the progress and conduct of trials.
- Prepared standard operating procedures for the Clinical Trial Information department in compliance with federal regulations.

AYERST LABORATORIES, New York, New York   1984-1985
Biostatistician
- Developed and modified SAS programs for statistical analysis and summary data presentation.
- Coordinated production efforts with coders, programmers, and monitors in the design of data files and data displays.
- Participated in the design of clinical trials, case report forms and protocols.
- Reviewed medical monitor and medical writer's standard reports.

REVLON HEALTH CARE GROUP, Tuckahoe, New York   1983-1984
Assistant Biostatistician
- Performed statistical analyses of clinical data using SAS.
- Summarized and interpreted results of statistical analysis and findings.
- Prepared detailed reports including summary tables and corresponding figures.

PUBLICATIONS:

Sherwood, J., Sony, A., Salama, G., Obenauer-Kutner, L., Huyck, S., et al. Particle Size Coarsening Induced by Valve Silicone in a Metered Dose Inhaler. *Drug Development and Industrial Pharmacy, 33(2),* 155-162.