Published in final edited form as:

Environmetrics. 2013 May 1; 24(3): . doi:10.1002/env.2198.

A HIERARCHICAL FUNCTIONAL DATA ANALYTIC APPROACH FOR ANALYZING PHYSIOLOGICALLY BASED PHARMACOKINETIC MODELS

Siddhartha Mandal,

Norwegian Institute of Public Health, Oslo 0473, Norway, siddharta.mandal@fhi.no

Pranab K. Sen, and

University of North Carolina, Chapel Hill, NC 27599, pksen@bios.unc.edu

Shyamal D. Peddada

National Institute for Environmental Health Sciences, peddada@niehs.nih.gov

Abstract

Ordinary differential equation (ODE) based models find application in a wide variety of biological and physiological phenomena. For instance, they arise in the description of gene regulatory networks, study of viral dynamics and other infectious diseases, etc. In the field of toxicology, they are used in physiologically based pharmacokinetic (PBPK) models for describing absorption, distribution, metabolism and excretion (ADME) of a chemical in-vivo. Knowledge about the model parameters is important for understanding the mechanism of action of a chemical and are often estimated using non-linear least squares methodology. However, there are several challenges associated with the usual methodology. Using functional data analytic methodology, in this article we develop a general framework for drawing inferences on parameters in models described by a system of differential equations. The proposed methodology takes into account variability between and within experimental units. The performance of the proposed methodology is evaluated using a simulation study and data obtained from a benzene inhalation study. We also describe a R-based software developed towards this purpose.

Keywords

benzene kinetics; functional basis; differential equations; cubic splines

1. Introduction

Systems of ordinary differential equation (ODE) are often used to describe biological and physiological phenomenon. A few areas of application include gene regulatory networks, viral dynamics, immunology and toxicology. For example, in gene regulatory networks, ODE models are used to describe changes in concentration of mRNA and/or associated proteins (Polynikis et al., 2009). In case of a viral infection, one may study its dynamics by observing the population of naive cells, infected cells or viral load and their changes over time (Perelson, 2002). For this article we shall focus on toxicological applications, where ODE models are used to model chemical kinetics, or the flow of chemicals in the human or

animal body. Although the main application is toxicology, the methodology developed here can be adapted to other contexts such as those described above.

There is considerable interest among toxicologists to conduct mechanistic studies to understand the rate at which a chemical is absorbed, distributed, metabolized and excreted (ADME) in various parts (i.e. compartments) of the body such as stomach, blood, liver, kidney etc. Often these processes are described using a system of differential equations known as physiologically based pharmacokinetic (PBPK) models (Reddy et al. 2005). For example, O'Flaherty et al. (2001) described a PBPK model for Chromium VI, Gentry et al. (2004) described one for arsenic and Kawahara et al. (1999) for the drug digoxin. Existing strategies to analyze PBPK models involve non-linear least squares methodology that solves the differential equations numerically to minimize the residual sum of squares. However as noted in an international workshop entitled "The International Workshop on Uncertainty and Variability in PBPK Models" conducted by the US Environmental Protection Agency (EPA) in 2006, these strategies lack a formal statistical framework that account for uncertainty, variability, model inadequacy among other issues. The need for research in this area was highlighted in a publication resulting from the workshop Barton et al. (2007). In the last two decades, a non-linear mixed effects methodology has been used in the literature for differential equation models (Davidian and Giltinan, 1995; Tornøe et al., 2004). These methods assume a hierarchical distributional framework for the response, parameters and random effects. However, the solution to the system of equations is still essential and very often a Taylor series expansion is employed to obtain a first order approximation. In cases where a closed form solution is available these methods are expected to work well. Also these methods require numerically integrating out the random effects for each individual, which may turn out to be computationally intensive for large number of individuals.

Motivated by Barton et al. (2007), in this article we provide a statistically rigorous framework for analyzing PBPK models by adopting the functional data analytic approach developed in Ramsay (1996) and Ramsay et al. (2007). Ramsay (1996) developed Principal Differential Analysis (PDA) to analyze data under models defined by system of differential equations. The pivotal idea was to avoid the solution to the system of differential equations and instead model the state variables using a functional basis consisting of polynomials or cubic splines while regularizing the approximations by penalizing higher order derivatives. Ramsay et al. (2007), used the system of differential equations as penalty while carrying out the Principal Differential Analysis. Recent research on differential equation models in areas such as HIV dynamics and ecological data have used the generalized profiling approach by Ramsay et al. (2007). Three such instances are Liang and Wu (2008), Varziri et al. (2008) and Cao et al. (2011). However, there exist some limitations associated with these approaches, such as the need for all variables to be observable or difficulty in estimation of variance components. We extend the methodology in Ramsay (1996) and Ramsay et al. (2007) to conduct formal statistical inferences by taking into account potential intra and inter-individual variability and correlations between and within compartments. In Table 1, we summarize the main features of the above discussed papers and our contribution.

We describe a benzene inhalation study conducted by Yu (1995) in Section 2 along with the PBPK model, as an illustrative example to our work. In Section 3 we describe the problem formulation and the estimation methodology. We illustrate the proposed methodology using a simulated data and the benzene inhalation study in Section 4. Concluding remarks are provided in Section 5.

2. A Benzene Study

Benzene is a known carcinogen to humans and animals, with both occupational and non-occupational exposure. Savitz and Andrews (1997) shows evidence of increased cancer incidence due to exposure to benzene and its metabolites. Kinetics of benzene has been modeled using different PBPK models (Travis et al., 1990; Woodruff and Bois, 1993). They are compartmental models incorporating main tissues and blood acting as the mode of transport between these tissues. No direct exchange of benzene occurs between compartments and metabolites are formed within bone marrow and liver, the metabolizing compartments. The five tissue groups for the model we are considering are (1) Richly perfused tissues, (2) Slowly perfused tissues, (3) Fat, (4) Liver and (5) Bone marrow. A schematic representation of one such model is shown in Figure 1.

We now describe the system of differential equations that defines the PBPK model in Figure 1. In the PBPK equations, A_j denotes the amount of the chemical in the j -th compartment and Cv_j is the concentration of the chemical in the venous blood at the j -th tissue exit. Q_j denotes blood flow rate, V_j is the volume and p_j is the partition coefficient (ratio of concentrations of chemical in tissue and blood at equilibrium) for the j -th tissue. Q_t and Q_{alv} are the total blood flow and the alveolar blood flow respectively. P_b denotes the blood:air partition coefficient. These are the physiological parameters in the model for which the values are well established and taken from literature (Travis et al., 1990). V_{max} denotes the maximum metabolism rate and k_m is the concentration at which reaction rate is half of V_{max} for metabolizing sites. These constitute the unknown parameters in the model. C_{art} , C_{ven} , C_{in} and C_{exh} represent the concentration of benzene in arterial blood, venous blood, inhaled and exhaled breath, respectively.

In the non-metabolizing sites, namely richly perfused tissues (tissues with high blood flow, abbreviated as *rpt*), slowly perfused tissues (tissues with low blood flow, abbreviated as *spt*) and *fat*, the rate of change in amount of benzene is directly proportional to the instantaneous concentration in the respective tissue. Blood flow rate serves as the rate constant. Thus the ODE, for these compartments are given by the following:

$$\frac{dA_j}{dt} = Q_j (C_{art} - Cv_j), \quad j = \{rpt, spt, fat\}. \quad (2.1)$$

In bone marrow (abbreviated as *bm*) and liver (abbreviated as *liv*), there is an extra term describing non-linear Michaelis-Menten kinetics for the formation of metabolites, apart from the usual first order kinetics. The equations for these sites and the metabolites (abbreviated as *met*) are as follows:

$$\frac{dA_j}{dt} = Q_j (C_{art} - Cv_j) - \frac{V_{max(j)}Cv_j}{k_m(j) + Cv_j}, \quad j = \{liv, bm\}. \quad (2.2)$$

$$\frac{dA_{met}}{dt} = \frac{V_{max(bm)}Cv_{bm}}{k_m(bm) + Cv_{bm}} + \frac{V_{max(liv)}Cv_{liv}}{k_m(liv) + Cv_{liv}}. \quad (2.3)$$

The concentrations in venous blood, arterial blood and exhaled breath are expressed as

$$C_{ven} = \sum_{j \in \{rpt, spt, fat, liv, bm\}} \frac{Q_j Cv_j}{Q_t}. \quad (2.4)$$

$$C_{art} = \frac{Q_t C_{ven} + Q_{alv} C_{in}}{Q_t + \frac{Q_{alv}}{P_b}}. \quad (2.5)$$

$$C_{exh} = \frac{C_{art}}{P_b}. \quad (2.6)$$

And the venous blood concentrations at the j -th tissue exit is given by

$$C_{vj} = \frac{A_j}{V_j p_j}. \quad (2.7)$$

The above equations (2.1)-(2.7) represent the PBPK model for benzene inhalation shown in Figure 1.

Yu (1995) conducted a benzene inhalation study where four individuals were exposed to benzene through inhaled air for two hours at low concentrations, after which the exposure was stopped. The subjects differed in terms of body weights and exposure concentrations. Benzene concentration in the exhaled breath of the subjects was measured at 5, 15, 30, 60, 90, 120 and 150 minutes after exposure was stopped. The observed concentrations in micrograms per cubic meter for the four individuals are provided in Table 2. This dataset is used as an illustrative example for our developed methodology.

3. The Proposed Methodology

Let p denote the total number of states in a PBPK model. Suppose only d out of p compartments in the system are measured for the i -th subject, $i = 1, 2, \dots, n$. Suppose

$\mathbf{Z}_i = (Z_{i1}, Z_{i2}, \dots, Z_{id}, Z_{i(d+1)}, \dots, Z_{ip})' = (\mathbf{Z}_i^o, \mathbf{Z}_i^u)'$ denotes the vector of state variables in the p states, where $\mathbf{Z}_i^o$ denotes the observable and $\mathbf{Z}_i^u$ denotes the unobservable part for the i -th subject. Let the system of differential equations that describe the rate of change in $\mathbf{Z}_i$ be given by:

$$\dot{\mathbf{Z}}_i = \mathbf{F}(\mathbf{Z}_i, t, \theta_i) \quad (3.1)$$

where $\dot{\mathbf{Z}}_i = d\mathbf{Z}_i/dt$ denotes the rate of change in $\mathbf{Z}_i$, $\mathbf{F}(\mathbf{Z}_i, t, \theta_i)$ is a $p \times 1$ vector of known functions, $\theta_i = (\theta_{i1}, \dots, \theta_{im})'$ is the vector of unknown parameters. Specifically, in the case of 2, $\mathbf{Z}$ represents the amount of benzene in the tissue the benzene study described in Section compartments, $\mathbf{F}(\cdot)$ represents the different linear and/or non-linear functions that describe the rates of change in amounts of benzene and are the metabolic parameters for liver and bone marrow compartments. We summarize the formulation in terms of state variables, functions and parameters for several applications in Table 3.

It is important to note that, due to experimental and other practical reasons, not all state variables are measured according to the same measurement protocol. Thus not all compartments are measured with the same frequency. For example, in rodent studies the frequency of drawing urine samples from a given mouse may be greater than the frequency of drawing blood samples. Let $(a_1, \dots, a_d)'$ be the number of measurement times for the d observable compartments. Let $\mathbf{Y}_i$ be the data for the observable component $\mathbf{Z}_i^o$ for the i -th subject. Thus we have, $\mathbf{Y}_i = \mathbf{Z}_i^o + \mathbf{e}_i$, where the random error $\mathbf{e}_i = (e_{i1}, \dots, e_{id})'$ is a $(a_1 + \dots + a_d) \times 1$ vector. Note that in some experimental settings, the observed variable may not be concentrations from any particular compartment(s) in the model. For example, in the

benzene PBPK model, exhaled breath concentration (observable) is not a compartment in the model. Rather it is a combination of all the five compartments. In such cases $Y_i = g(Z_i^o) + e_i$, where $g(\cdot)$ is a known function from the specified model.

One may assume that e_{ij} are independently distributed for all i with mean $\mathbf{0}$ and $a_j \times a_j$ covariance matrix $\Omega_{ij}(t)$. It is reasonable to assume that the intra-individual correlation between measurements would decrease with amount of separation in time. Hence as commonly used in time-series models for a parsimonious description of dependence in data (Box et al. 2008), we let

$$\text{Corr}(e_{ij}(t_1), e_{ij'}(t_2)) = \begin{cases} \rho_{i(j)}^{|t_1-t_2|}, & j=j' \\ \rho_{i(j,j')}^{|t_1-t_2|}, & j \neq j'. \end{cases}$$

Lastly we model inter-individual variability by imposing a hierarchical structure on θ_i . We shall assume that θ_i are identically and independently distributed with mean θ and covariance matrix Ψ , which estimates the inter-individual variability.

3.1. Individual parameter estimation

We begin by estimating parameters for each individual subject. The functional basis approach of Ramsay et al. (2007) is used to approximate the state variables Z_i in (3.1). This amounts to selecting a class of basis functions, such as splines or polynomials, to approximate Z_i . In this article we use B-splines as the basis functions. Let

$\tilde{Z}_{ij}(t)$, $j=1, \dots, p$ be the approximate value of compartment j at time t , described through L_j basis functions. Then $\tilde{Z}_{ij}(t) = \sum_{l=1}^{L_j} \phi_{ijl}(t) \alpha_{ijl} = \phi_{ij}(t)' \alpha_{ij}$, where $\phi_{ij}(t)$ is a vector of basis functions and α_{ij} is a vector of unknown basis parameters. Let $\alpha_i = (\alpha_{i1}, \dots, \alpha_{ip})'$ and $\phi_i(t) = (\phi_{i1}'(t), \dots, \phi_{ip}'(t))$. Hence the approximated state variable for the i -th subject is $\tilde{Z}_i(t) = (\tilde{Z}_{i1}(t), \dots, \tilde{Z}_{ip}(t))' = \phi_i(t) \alpha_i$ and $\tilde{Z}_i = \Phi_i \alpha_i$ where $\Phi_i = \text{diag}(\phi_i(t_1), \dots, \phi_i(t_q))$ and $t_1, \dots, t_q$ are the measurement times. The basis parameter vector α_i are determined such that $\tilde{Z}_i^o$ closely mimics the data and the state variables $\tilde{Z}_i$ satisfies the system of differential equations. Hence both $\tilde{Z}_i$ and $\tilde{Z}_i^o$ are functions of the nuisance parameters α_i .

We follow a two-step procedure to estimate the individual model parameters θ_i by obtaining the basis approximation based on the data and differential equations and then use the estimated approximation to minimize a predicted residual sum of square. For a fixed θ_i , we estimate α_i and ρ_i such that the “distance” between the observed Y_i and the approximation of the corresponding observable $\tilde{Z}_i^o$ is minimized subject to the underlying differential equations. Thus, for a given θ_i , we minimize $H(\alpha_i, \rho_i, \lambda_i)$ with respect to α_i , ρ_i and λ_i :

$$H(\alpha_i, \rho_i, \lambda_i) = (Y_i - \tilde{Z}_i^o(\alpha_i))' \{\Omega(\rho_i)\}^{-1} (Y_i - \tilde{Z}_i^o(\alpha_i)) + \lambda_i f(\dot{\Phi}_i \alpha_i - F(\Phi_i \alpha_i, t, \theta_i))' (\dot{\Phi}_i \alpha_i - F(\Phi_i \alpha_i, t, \theta_i)) \quad (3.2)$$

In the above expression λ_i is a regularization parameter which is estimated in the above objective function. The estimators $\hat{\alpha}_i$, $\hat{\rho}_i$ and $\hat{\lambda}_i$ obtained by minimizing (3.2) are implicit functions of θ_i . Hence the of estimated state variables $\hat{\tilde{Z}}_i$ and therefore $\hat{\tilde{Z}}_i^o$ are implicit functions of θ_i . For notational simplicity, we denote the predicted value of the observable

state variable by $\hat{\mathbf{Z}}_i^o(\theta_i)$. Using the estimators derived by minimizing (3.2), we minimize the following quadratic form to estimate the model parameter θ_i :

$$S(\theta_i) = \left(\mathbf{Y}_i - \hat{\mathbf{Z}}_i^o(\theta_i) \right)' \{ \Omega(\hat{\rho}_i(\theta_i)) \}^{-1} \left(\mathbf{Y}_i - \hat{\mathbf{Z}}_i^o(\theta_i) \right) \quad (3.3)$$

The objective functions (3.2) and (3.3) are iteratively optimized until convergence. Using the delta method for $\hat{\theta}_i$, we obtain an approximate covariance matrix of $\hat{\theta}_i|\theta_i$ given by

$$\left(\frac{d\hat{\theta}_i}{d\mathbf{Z}_i^o} \right) \Omega(\hat{\rho}_i) \left(\frac{d\hat{\theta}_i}{d\mathbf{Z}_i^o} \right)', \text{ henceforth denoted by } \Gamma(\theta_i) \text{ in this paper.}$$

3.2. Population parameter estimation

In order to estimate the population parameters using the estimates obtained from each subject, we assume the following hierarchical model structure for $\hat{\theta}_i$ and θ_i .

Assumptions A:

(A.1) $\hat{\theta}_i|\theta_i$ independently distributed with mean θ_i and covariance $\Gamma(\theta_i)$.

(A.2) $\theta_i|\theta$ i.i.d with mean θ and covariance Ψ .

Then, $E(\hat{\theta}_i) = E_\theta(E(\hat{\theta}_i|\theta)) = \theta$ and $Cov(\hat{\theta}_i) = \Psi + E_\theta(\Gamma(\theta_i)) = \mathbf{V}_\theta$. This is a Type III non-linear marginal model described in Demidenko (2005). Our objective is to estimate the population parameter θ and the covariance matrix of $\hat{\theta}$. Since $\mathbf{V}_\theta$ is a function of θ , the classical iterated weighted least squares type methodology is not applicable here. Demidenko (2005) suggests a Total Generalized Estimating equation approach for such a formulation. However TGEE cannot be implemented in the present problem since the explicit form of $\mathbf{V}_\theta$ is unknown. Hence we use an empirical Bayes based technique to solve the problem along the lines of Zhang et al. (2000). We start with initial estimates of θ and Ψ , obtained from

$$\hat{\theta}_{(0)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i, \quad \hat{\Psi}_{(0)} = \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \hat{\theta}_{(0)}) (\hat{\theta}_i - \hat{\theta}_{(0)})'$$

and iterate between the two following steps:

Step 1: Update $\hat{\theta}_i$ as

$$\hat{\theta}_{i,(c+1)} = \left(\hat{\Gamma}_i^{-1} + \hat{\Psi}_{(c)}^{-1} \right)^{-1} \left(\hat{\Gamma}_i^{-1} \hat{\theta}_{i,(c)} + \hat{\Psi}_{(c)}^{-1} \hat{\theta}_{(c)} \right),$$

where $\hat{\Psi}_{(c)}$ and $\hat{\theta}_{(c)}$ are the estimates from the c -th iterate.

Step 2: Update the population parameter estimates as

$$\hat{\theta}_{(c+1)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{i,(c+1)}, \quad \hat{\Psi}_{(c+1)} = \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_{i,(c+1)} - \hat{\theta}_{(c+1)}) (\hat{\theta}_{i,(c+1)} - \hat{\theta}_{(c+1)})'$$

Let $\hat{\Psi}$ and $\hat{\theta}$ denote the estimates upon convergence and $\hat{\mathbf{V}}_{\theta} = \hat{\Psi} + \sum_{i=1}^n \Gamma(\hat{\theta}_i)$.

4. Simulated and Real Data Examples

In this section we present simulated and real data based on the PBPK model of benzene described earlier. The simulated example is designed according to the real data example used in this paper. We simulated a random sample of five subjects such that each subject is exposed to $161 \mu\text{g}/\text{m}^3$ benzene through inhalation for two hours, following which the subjects were provided with clean breathing air. Concentration of benzene in exhaled breath at 5, 15, 30, 60, 90, 120 and 150 minutes post-exposure is the response variable for each individual. The PBPK model described in Section 2 is used as the model of interest. We obtain population parameter estimates of the metabolic parameters along with estimates of their variability.

We treat the maximum metabolic rates in liver and bone marrow as unknown parameters. So $\theta = (V_{\max(\text{liver})}, V_{\max(\text{bm})}, \log(k_{\text{m}(\text{liver})}), \log(k_{\text{m}(\text{bm})}))'$ and its true population value is taken as $\theta = (387.4, 79.2, 0.69, 2.71)'$. We assume that the individual parameter values $\theta_i \sim \text{iid} N_4(\theta, \mathbf{J} + \text{diag}(50, 3, .01, .1))$, where $\mathbf{J}$ is a matrix of 1's. Concentration in exhaled breath (X_i) is obtained by solving the benzene PBPK model equations given by (2.1)-(2.7) using θ_i . Finally the data Y_i is generated from $N_7(X_i, \mathbf{R})$. Here $\mathbf{R}$ is an intra-individual covariance

structure. We choose the (t, t') -th element of $\mathbf{R}$ as $r_{t,t'} = 5(0.2)^{|t-t'|}$. We look at 100 datasets consisting of five individuals each. Linear combinations of nine B-splines of order 4 are used to approximate the concentration of benzene in each compartment and the metabolites concentration. Simulated annealing is applied to obtain the basis and model parameter estimates for each individual. We use the iterative methodology described in Section 3.2 to obtain the population parameter estimates and the corresponding variability estimates.

4.1. Results of Simulated Example

Results of the simulation study, summarized in Table 4, suggest that the proposed methodology has a small relative bias and relative MSE (relative to true value of the parameter) for a sample size as small as five. For one of the parameters, $k_{\text{m}(\text{liver})}$ the estimate fails to capture the true value and the relative MSE and bias are enlarged due to the small true parameter value. The estimated coverage probability for a 95% joint confidence region of θ centered at $\hat{\theta}$ is 0.92 (s.e. = 0.027). This suggests that even with a sample size as small as 5 the proposed methodology yields reasonably accurate confidence regions for the population parameter.

4.2. Benzene Inhalation Study

We applied the proposed methodology to the benzene data described in Section 2. The unknown population parameter of interest θ is $(V_{\max(\text{liver})}, V_{\max(\text{bm})}, k(k_{\text{m}(\text{liver})}), k(k_{\text{m}(\text{bm})}))'$. For each individual, we approximated the five compartments and metabolites using nine B-spline basis functions of order 4. Using these approximations we obtained the estimates $\hat{\theta}_i$, $i = 1, \dots, 4$. Solving the benzene PBPK model equations with $\hat{\theta}_i$ as the input parameter estimate, we obtained the predicted benzene concentration in breath of each of the four subjects. The individual fitted curves presented in Figure 2 suggest that the proposed methodology fits the observed data very well visually.

The estimated values of the intra-individual correlation coefficients $\hat{\rho}_i$ for the four subjects are 0.146, 0.2434, 0.623 and 0.025. Estimated values of λ for the four individuals are 2.38×10^{-5} , 1.98×10^{-4} , 1.06×10^{-4} and 1.37×10^{-6} . The population parameter estimates obtained

are as follows: $V_{\max(liv)} = 387.41$, $V_{\max(bm)} = 79.25$, $k_{\max(liv)} = 1.22$, $k_{\max(liv)} = 17.04$. The estimated total variability is

$$\hat{V}_{\theta} = \begin{pmatrix} 20556.87 & 4957.96 & -307.30 & 471.16 \\ 4957.96 & 1292.01 & -78.07 & 77.02 \\ -307.30 & -78.07 & 4.77 & -7.08 \\ 471.16 & 77.02 & -7.08 & 190.22 \end{pmatrix}.$$

Using the population estimate $\hat{\theta}_i$ in Figure 3 we plot the exhaled benzene concentration of a person who was exposed to average concentration of the four subjects in the sample, superimposed with the observed curves for the 4 individuals.

The population parameter estimate $\hat{\theta}$ obtained above can also be used for predicting the exhaled benzene concentration of a randomly chosen individual from the population by solving the differential equations in (2.1)-(2.7). For example, in Figure 4 we provide the predicted exhaled benzene concentration of a randomly chosen person who weighs 130 lbs (average of the four individual body weights) and who was exposed to $161 \mu\text{g}/\text{m}^3$ (average exposure concentration) benzene for two hours. The vertical bars are the 95% prediction intervals at various time points.

In Figure 5 we analyze the predicted behavior of benzene as explained by the system of differential equations. We consider the post exposure concentration of benzene in the different compartments of the PBPK model and exhaled breath across time with the solution of the PBPK model using the estimated population parameters. As in any other context, the performance of a fitted model can only be judged well within the range of the observed data. Extrapolation outside the observed range of the data needs to be interpreted with caution. This holds even in the present situation.

We investigate a few important features of the predicted model. For all the compartments and exhaled breath, the concentration reaches a peak at 120 minutes and decreases post exposure. The plots give us an idea as to how and where the benzene is being processed. Both fat and slowly perfused compartments show a slow decrease in concentration than others, indicating an affinity of benzene towards these classes of tissues. As for the metabolizing sites, liver and bone marrow, the concentrations decrease rapidly. This could indicate formation of metabolites of benzene in these two sites. Also the rate of metabolism appears to be faster in liver than in bone marrow. The information here provides an insight into the kinetic behavior of benzene which was one of our main objectives.

4.3. Software and Programming

All the computation and programming was done using the R programming language (version 2.15.0). We have used some pre-existing R packages (fda, Bolstad2, maxLik, mnormt, MASS) and developed a user friendly interface that can be used to perform individual parameter estimation for repeated measurement data using the PBPK model for benzene. One can change the R source code in order to analyze other differential equations. We are currently looking into ways such that it can be accomplished through the user interface itself.

5. Conclusion

Differential equation modeling is an important area of research with several applications in fields such as engineering, biology and toxicology, especially in context of pharmacokinetic

and toxicokinetic modeling. Although such models are widely used in a variety of contexts, as noted in the recent EPA workshop and the resulting publication (Barton et al., 2007) there does not exist a well-developed statistical methodology for drawing inference on the model parameters. This research, which exploits the functional data analytic approach of Ramsay et al. (2007), takes the first step towards a formal statistical theory and methodology. Specifically, an important contribution of the proposed methodology is that it accounts for; (i) inter and intraindividual variability, (ii) the dependence within subjects and between compartments/states. Secondly, our methodology overcomes the computational burden of usual strategies by avoiding the problem of numerically solving a system of differential equations. This in turn also alleviates situations where the differential equations explain the behavior of the chemical better in some compartments than others. The PBPK model for benzene used in this paper is a general model for a volatile compound exposed through inhalation. Although the choice of the model would depend upon the chemical and the route of exposure, our statistical formulation and approach would remain valid.

In animal studies, it is not uncommon to obtain data on various compartments with different frequencies. In ongoing research we are extending the proposed methodology to accommodate such unequal time spacing. Toxicologists are often interested in a variety of comparisons involving the model parameters. For example, in animal studies one may want to investigate differences in model parameters among sex, species or strain of animals. The methodology developed in this paper can be extended to make such comparisons by incorporating suitable covariates (continuous or categorical) in the model.

Acknowledgments

The authors would like to thank Dr. Grace Kissling and Dr. Amy Herring for reading the manuscript and providing valuable inputs and suggestions.

FUNDING INFORMATION

This work was supported in part by Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences [Z01 ES101744-04].

REFERENCES

- Barton H, Chiu W, Woodrow S, Andersen M, Bailer A, Bois F, Dewoskin R, Hays S, Johanson G, Jones G, Loizou N, Macphail R, Portier C, Spendiff M, Tan Y. Characterizing uncertainty and variability in physiologically based pharmacokinetic models: state of the science and needs for research and implementation. *Toxicological Sciences*. 2007; 99:395–402. [PubMed: 17483121]
- Box, G.; Jenkins, G.; Reinsel, G. Time series analysis: forecasting and control. John Wiley; Hoboken, N.J.: 2008.
- Cao J, Wang L, Xu J. Robust Estimation for Ordinary Differential Equation Models. *Biometrics*. 2011; 67:1305–1313. [PubMed: 21401565]
- Davidian, M.; Giltinan, D. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman and Hall; USA: 1995. Nonlinear Models for Repeated Measurement Data.
- Demidenko, E. Mixed Models: Theory and Applications. John Wiley & Sons, Inc.; Hoboken, NJ, USA: 2005. Nonlinear Marginal Model.
- Gentry P, Covington T, Mann S, Shipp A, Yager J, Clewell HI. Physiologically based pharmacokinetic modeling of arsenic in the mouse. *Journal of Toxicology and Environmental Health, Part A*. 2004; 67:43–71. [PubMed: 14668111]
- Kawahara M, Sakata A, Miyashita T, Tamai I, Tsuji A. Physiologically based pharmacokinetics of digoxin in mdr1a knockout mice. *Journal of Pharmaceutical Sciences*. 1999; 88:1281–1287. [PubMed: 10585223]

- Liang H, Wu H. Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of the American Statistical Association*. 2008; 103:1570–1583. [PubMed: 19956350]
- O’Flaherty EJ, Kerger B, Hays S, Paustenbach D. A physiologically based model for the ingestion of chromium(iii) and chromium(vi) by humans. *Toxicological Sciences*. 2001; 60:196–213. [PubMed: 11248132]
- Perelson A. Modeling viral and immune system dynamics. *Nature Reviews Immunology*. 2002; 2:28–36.
- Polynikis A, Hogan S, Bernardo M. Comparing different ode modeling approaches for gene regulatory networks. *Journal of Theoretical Biology*. 2009; 261:511–530. [PubMed: 19665034]
- Ramsay J, Hooker G, Campbell D, Cao J. Parameter estimation for differential equations: A generalized smoothing approach. *Journal of the Royal Statistical Society, Series B*. 2007; 69(5): 741–796.
- Ramsay JO. Principal differential analysis: Data reduction by differential operators. *Journal of the Royal Statistical Society, Series B*. 1996; 58:495–508.
- Reddy, M.; Yang, R.; Andersen, M.; Clewell, H. *Physiologically Based Pharmacokinetic Modeling: Science and Applications*. Wiley and Sons; New York, USA: 2005.
- Savitz D, Andrews K. Review of epidemiologic evidence on benzene and lymphatic and hematopoietic cancers. *American Journal of Industrial Medicine*. 1997; 31(3):287–295. [PubMed: 9055951]
- Tornøe C, Agersø H, Jonsson E, Madsen H, Nielsen H. Non-linear mixed-effects pharmacokinetic/pharmacodynamic modelling in nlme using differential equations. *Computer methods and program in biomedicine*. 2004; 76:31–40.
- Travis C, Quillen J, Arms A. Pharmacokinetics of benzene. *Toxicology and Applied Pharmacology*. 1990; 102:400–420. [PubMed: 2315912]
- Varziri M, McAuley K, McLellan P. Parameter estimation in continuous-time dynamic models in the presence of unmeasured states and nonstationary disturbance. *Ind. Eng. Chem. Res*. 2008; 47:380–393.
- Woodruff T, Bois F. Optimization issues in physiological toxicokinetic modeling: a case study with benzene. *Toxicology Letters*. 1993; 69:181–196. [PubMed: 8212060]
- Yu, R.; Rutgers University. Ph.D. thesis. 1995. Measurement of Partitioning of Benzene in Breath and Urinary Metabolites from Benzene Exposure in Humans.
- Zhang, J.; Peddada, S.; Rogol, A. Estimation of parameters in nonlinear regression models. In: Rao, CR.; Szekely, G., editors. *Statistics for the 21st Century*. 2000. p. 459-483.

Figure 1.

A five compartment Benzene PBPK model. The boxes represent the compartments and the dashed arrows represent the flow of chemical through blood. It is a schematic representation of the tissues in human physiology that are related to benzene kinetics.

Figure 2.

Estimated curves (solid black lines) for the four individuals obtained by solving the differential equations with the estimated individual parameter values. Symbols for individual data points are (■).

Figure 3. Population estimated curve with data for four individuals. The solid black line shows the exhaled breath concentration obtained from the solution of the Benzene PBPK model using the population model parameter estimates.

Figure 4. Population fitted curve (dashed line) with 95% prediction intervals (solid vertical lines). The population fitted curve is obtained from the solution of the system of differential equations using the estimated population parameters. The prediction intervals are obtained through the delta method.

Figure 5.

Compartmental concentrations (in units of micrograms per cubic meter) across time. These plots are obtained by solving the differential equations defining the benzene PBPK model using the estimated population model parameter estimates.

Table 1

Main features of some recent papers based on Ramsay et al. (2007) compared to the proposed methodology.

Feature	Varziri et al. (2008)	Liang and Wu (2008)	Proposed methodology
1. Allows unobservable compartments	Yes	No	Yes
2. Basis approximation uses the data and the differential equations.	Yes	No	Yes
3. Subject specific inference.	No	Yes	Yes
4. Population based inference.	Yes	No	Yes
5. Separate estimation of variance components.	No	No	Yes

Table 2

Exhaled breath concentrations (in $\mu\text{g}/\text{m}^3$) from the benzene inhalation experiment on the four subjects over seven time points after exposure was stopped at two hours (Yu, 1995).

Time post-exposure	5 mins	15 mins	30 mins	60 mins	90 mins	120 mins	150 mins
Subject 1	12.0	4.75	4.6	3.9	1.9	1.2	1.7
Subject 2	9.2	5.2	3.9	2.95	2.5	2.1	1.8
Subject 3	5.1	3.5	3.9	2.0	1.6	0.083	1.5
Subject 4	8.8	6.65	4.7	3.9	3.2	2.1	2.6

Table 3

Some examples of applications of the proposed methodology.

	State variables	Functions	Parameters
Gene-regulatory networks	mRNA concentrations or gene expressions	Factors of translation, transcription, degradation etc.	Rates of the factors affecting concentration or expression
Viral dynamics	Density of naive cells, infected cells, dead cells	Factors of infection, death etc.	Rates of viral infection, susceptibility etc.
Immunology	Proportion of centroblasts, centrocytes	Factors of mutation, selection, proliferation	Rates of mutation, selection etc.
Toxicology	Amounts or concentration of chemicals	Factors of chemical processes, eg: absorption, metabolism	Metabolic parameters, Rates of reaction

Table 4

Population parameter estimation results for the simulated data example.

	$V_{\max(liv)}$	$V_{\max(bm)}$	$\log(k_{m(liv)})$	$\log(k_{m(bm)})$
Units	$\mu\text{g}/\text{min}$	$\mu\text{g}/\text{min}$	$\mu\text{g}/\text{l}$	$\mu\text{g}/\text{l}$
True value	387.4	79.2	0.69	2.71
Estimate (Mean)	386.97	80.02	1.21	2.82
Rel MSE(%)	0.05	0.84	39.03	0.49
Rel Bias(%)	-0.11	1.03	75.03	4.23