



Statistical Analysis of PBPK Models

R. Garcia¹, E. Kenyon², and R. W. Setzer³

¹UNC Chapel Hill, Dept. of Biostatistics; ²National Health and Environmental Effects Research Laboratory, US EPA, RTP, NC; ³National Center for Computational Toxicology, US EPA, RTP, NC US EPA, ORD, Computational Toxicology Research Program

research development

Project ID: IIID-1

Science Question

One of the reasons statistical analysis of physiologically based pharmacokinetic (PBPK) models can be difficult is that such models usually contain a large number of parameters relative to the data available to estimate them. While information is generally available about some of these parameters prior to data analysis, there is relatively little or no prior information available for other parameters. When the available data do not provide information on a parameter for which there is also little prior information then this parameter is known as unidentifiable. Unidentifiable parameters are problematic because they can lead to problems with numerical procedures used in parameter estimation. In addition, some estimation procedures may give misleading results when parameters are unidentifiable.

Research Goals

Parameter identifiability is examined for simple types of PBPK models and these results are generalized to more common and complex types of PBPK models. PBPK models considered are

- homogeneous and nonhomogeneous first order linear differential equations. A PBPK model of dimethylarsenic in rats (Kenyon, 2001) was used as a concrete example.
- non-homogeneous nonlinear first order differential equations.

This research attempts to:

- establish a list of conditions that would determine whether parameters or functions of parameters are identifiable;
- make suggestions on how to amend a PBPK model when unidentifiable parameters are encountered;
- determine how to specify appropriate prior distributions and likelihoods, in the presence of unidentifiable parameters, for Bayesian data analysis;
- explore approaches to using parallel computing to speed up statistical analyses of PBPK models.

Long Term Goal III

What Does "Identifiability" Mean?

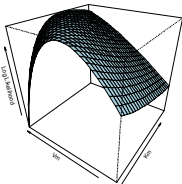
Two types of identifiability have been defined:

1. structural identifiability
2. statistical identifiability

Structural non-identifiability of parameters in PBPK models refers to parameters or function of parameters which are not estimable due to the model formulation. Structural non-identifiability persists regardless of the amount of data present. Statistical non-identifiability occurs when lack of appropriate data prevents estimation of parameters.

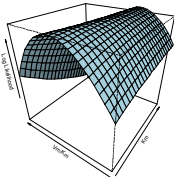
As an example to illustrate statistical identifiability, consider a single compartment model with Michaelis Menten elimination (Km=6, Vmax=9). If data concentrations are low relative to Km (concentrations < 5), then it will not be possible to individually estimate Vmax and Km for metabolic clearance. The plot below illustrates the log-likelihood function for different values of Km and Vm. The plot shows the uncertainty associated with both Km and Vm since a clear maximum does not exist.

Log Likelihood for different values of Km and Vm



Contrastingly, a plot of the likelihood function for values of Km and Vm/Km indicates that the ratio has a clear maximum, therefore it can be estimated. However, uncertainty associated with Km still remains. This indicates that Km and Vm are both individually statistically non-identifiable however their ratio is identifiable.

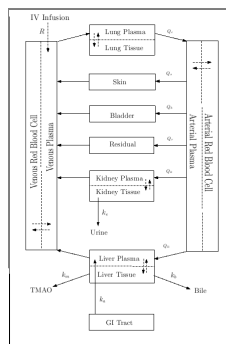
Log Likelihood values for different values of Km and Vm/Km



Methods/Approach

PBPK model of Dimethylarsenic in Rats.

This model can be expressed as non-homogenous system of linear differential equations.



Initially, conditions for identifiability of parameters in simple types of PBPK models are established and then these results are generalized to more common and complex types of PBPK models. The reason for examining simple models first is because the conditions necessary to establish identifiability in simple models will most likely also be necessary in more complex models.

First, parameter identifiability of homogeneous linear differential equations were examined. These include PBPK models in which all absorption, metabolism, and elimination rates can be described by zero or first order rates. Specifically, these models are such that the system of differential equations for the concentrations, $f(t, \theta)$, can be expressed as

$$\dot{x}(t, \theta) = A(\theta, x) f(t, \theta)$$

where the variables t and θ denote time and parameters respectively and the term $A(\theta, x)$ is a matrix which is a function of θ and independent variables, x . For these types of models a closed form solution to the system of differential equations exists,

$$f(t, \theta) = \exp(A(\theta, x)t) f(0)$$

where the term $f(0)$ denotes concentrations at time 0.

First, conditions are established for PBPK models in which A can be expressed as a linear function of θ . This can be performed for some PBPK models by reformulating the parameters in the model.

Under these assumptions, the vec operator of the A matrix allows the reexpression, $\text{vec}(A(\theta)) = C\theta$, where C is a matrix. Two conditions which imply structural identifiability of the parameter θ of the solution to system of differential equations are..

- the matrix norm (size) of A be small
- C be of full column rank.

The second property is analogous to the estimability condition required of design matrices in simple linear regression.

Advantages of performing Statistical Analyses in R

Statistical analyses were performed using the R statistical language. Some of the advantages of using R are,

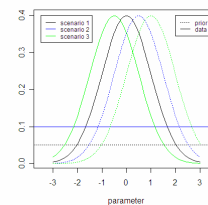
- allows for nearly any type of distributional assumption for the data
- natural framework for implementation of Markov Chain Monte Carlo techniques
- clustering of computations increases speed of analyses
- Flexible test platform for exploring parallel computing approaches (at what level is it most efficient to carry out the parallelization?)

In order for a parameter to be estimable, information about this parameter needs to be provided. This information can be from two sources,

- prior information from previous studies or the literature
- data through the likelihood specification

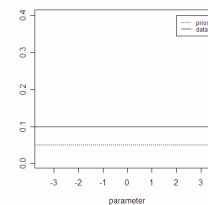
When information is being provided about a parameter then it is through one of the three scenarios displayed in the figure below.

3 Scenarios where information is being provided about parameter from prior or data



If no information is provided about a parameter through prior knowledge or data, see figure below, then this causes instability of numerical procedures to obtain estimates. In order to avoid this situation, one can first determine whether a parameter is statistically identifiable. If it is then this parameter is estimable since information is being provided about it through the data. If it is not statistically identifiable, then a prior can be constructed such that it does provide information about this parameter. This guarantees that at least some information is being provided about parameters. Three possible scenarios in which

Scenario in which NO information is being provided about parameter



Results/Conclusions

In PBPK models with homogeneous linear first order differential equations for which $A(\theta)$ is linear function of θ , the vec operator of the A matrix allows the reexpression, $\text{vec}(A(\theta)) = C\theta$, where C is a matrix. Two conditions which imply structural identifiability of the parameter θ of the solution to system of differential equations are

- 1) the norm of A be small
- 2) the matrix C be of full column rank.

Impact and Outcomes

Before PBPK models can be used in a robust way in risk assessments, the uncertainty of their predictions needs to be quantifiable. Currently, the best approach to that requires the use of Bayesian statistical methodology, to allow varying levels of information about parameters to be incorporated in the parameter estimation process. However, such estimates are unreliable when parameters are not sufficiently identified by the data, and the resulting failure of identifiability goes undiagnosed. This research will allow identifiability issues to be diagnosed and remedied, allowing more reliable estimation of parameters as well as more reliable quantification of the uncertainty of those estimates.

Future Directions

- Establish conditions needed to satisfy identifiability of PBPK models which can be expressed as a first order nonlinear non-homogenous system of equations;
- Determine what requirements are needed of prior distributions to establish a proper posterior;
- Begin to compare alternative levels of parallelization: at the level of individual chains, or within a chain, parallelizing the likelihood calculation.

References

Gelman, A., Bois, F. Y., and Jiang, J. (1996) Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *J. Am. Statist. Ass.*, 91, 1400-12

Gelfand, A. E. and Sahu, S. K. (1999) Identifiability, improper priors, and Gibbs sampling for generalized linear models. *J. Am. Statist. Ass.*, 94, 247-253.

Kenyon, E. M., Hughes, M. F. "A Concise Review of the Toxicity and Carcinogenicity of Dimethylarsenic Acid." *Toxicology* 160 (2001): 227-236.



This poster does not necessarily reflect EPA policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.