Physiologically based pharmacokinetic (PBPK) models are used in mode-of-action based risk and safety assessments to estimate internal dosimetry in animals and humans. When used in risk assessment, these models can provide a basis for extrapolating between species, doses, and exposure routes or for justifying nondefault values for uncertainty factors. Characterization of uncertainty and variability is increasingly recognized as important for risk assessment; this represents a continuing challenge for both PBPK modelers and users. Current practices show significant progress in specifying deterministic biological models and nondeterministic (often statistical) models, estimating parameters using diverse data sets from multiple sources, using them to make predictions, and characterizing uncertainty and variability of model parameters and predictions. The International Workshop on Uncertainty and Variability in PBPK Models, held 31 Oct–2 Nov 2006, identified the state-of-the-science, needed changes in practice and implementation, and research priorities. For the short term, these include (1) multidisciplinary teams to integrate deterministic and nondeterministic/statistical models; (2) broader use of sensitivity analyses, including for structural and global (rather than local) parameter changes; and (3) enhanced transparency and reproducibility through improved documentation of model structure(s), parameter values, sensitivity and other analyses, and supporting, discrepant, or excluded data. Longer-term needs include (1) theoretical and practical methodological improvements for nondeterministic/statistical modeling; (2) better methods for evaluating alternative model structures; (3) peer-reviewed databases of parameters and covariates, and their distributions; (4) expanded coverage of PBPK models across chemicals with different properties; and (5) training and reference materials, such as cases studies, bibliographies/glossaries, model repositories, and enhanced software. The multidisciplinary dialogue initiated by this Workshop will foster the collaboration, research, data collection, and training necessary to make characterizing uncertainty and variability a standard practice in PBPK modeling and risk assessment.

Key Words: physiologically based pharmacokinetic modeling; uncertainty; population variability; nonlinear modeling; risk assessment; Bayesian models.

Physiologically based pharmacokinetic (PBPK) models have been developed for numerous environmental and pharmaceutical compounds. Their uses include estimating internal doses (Bischoff et al., 1971; Reitz et al., 1980); supporting extrapolations across exposure routes (Clewell and Andersen, 1987), species (Clewell and Andersen, 1987), individuals (Mezzetti et al., 2003), age (Clewell et al., 2003); understanding the behavior of classes of compounds (Parham et al., 1997); reconstructing exposures (Georgopoulos et al., 1994); designing experiments (Bois et al., 1999); and testing hypotheses and theories (Andersen et al., 1987; Bois et al., 1991). Because of this broad utility, PBPK models are increasingly applied to risk assessments for...
chemical responses are associated with specific measures of tissue dose based on knowledge of or hypotheses for the biological processes (or mode of action) leading to toxicity. Chemical risk assessments use PBPK models for a range of purposes in the derivation of acceptable exposure levels (e.g., RfC, RfD) or cancer risk estimates, notably extrapolations across species, doses, and exposure routes or estimation of human pharmacokinetic variability for derivation of data-derived adjustments specific to the chemical of interest (U.S. EPA, 2006; WHO/IPCS, 2005). At the same time, in the risk management decisions informed by these models, there is a growing need for information on uncertainty—reflecting, in part, the level of confidence in model predictions—and variability—a reflection of the degree to which predictions may differ across a population. The International Workshop on Uncertainty and Variability in PBPK Models (UVPKM) was the first forum dedicated to a review and discussion of methodological and implementation issues in applying statistical methods to PBPK model-based analyses (31 October–2 November 2006 in Research Triangle Park, NC. Presentations, white papers, workgroup reports, and descriptions of follow-up activities are at www.epa.gov/ncct/uvpkm. Additional information is available at www.pbpk.org.). The sponsors of this international workshop seek to promote appropriate application of these methods by bringing together laboratory scientists, PBPK modelers, statisticians, and others with expertise in applied mathematics, biology, and pharmacokinetic analyses for cross-disciplinary exchanges of experience and ideas (Sponsors of the UVPKM: U.S. Environmental Protection Agency [EPA], Office of Research and Development, National Center for Environmental Assessment, National Center for Computational Toxicology, and National Health Effects and Environmental Research Laboratory. U.S. National Institutes of Health, National Institute of Environmental Health Sciences [NIEHS], Additional support was provided by: CIIT Centers for Health Research [CIIT], L’Institut National de l’Environnement Industriel et des Risques [INERIS], Miami University, Summit Toxicology, and U.K. Health and Safety Executive, Health and Safety Laboratory). As described in Table 1, the key scientific issues discussed were organized into the topical areas of model specification, calibration, and prediction. Because of the diverse expertise of the participants, background papers and presentations provided an overview of PBPK models, pharmacokinetic data, statistical models for uncertainty and variability, and risk assessment needs and applications (The invited speakers who contributed to this article through their presentations were: George M. Gray, EPA, Harvey J. Clewell III, CIIT; Hugh A. Barton, EPA; R. Woodrow Setzer, EPA; Weihsueh A. Chiu, EPA; Gunnar Johanson, Karolinska Institute; Melvin E. Andersen, CIIT; Marie Davidian, North Carolina State University (NCSU); and Frédéric Y. Bois, INERIS). In addition, three of the presenters prepared background papers prior to the meeting for review by the meeting participants: Harvey J. Clewell III, CIIT, Marie Davidian, NCSU, and Frédéric Y. Bois, INERIS). Workshop participants were then asked to

- Review and assess the status of existing methodologies for characterizing uncertainty and variability in PBPK models.
- Identify applications for which appropriate methodologies can be readily implemented given currently available data, information, and resources.
- Propose ways to improve/expand the implementation of these methods (including software and training).
- Identify key research priorities and/or data needs.

This article synthesizes highlights of the Workshop. We summarize the current state of the science on model specification, calibration, and prediction for risk assessment, recommend short-term improvements to current practice, and delineate medium- and long-term research and data needs that will greatly enhance the characterization of uncertainty and variability in PBPK models.

CURRENT PRACTICES IN MODEL DEVELOPMENT AND APPLICATION

Model Specification

Specification is the process for determining the structure and level of complexity needed in a particular model

<table>
<thead>
<tr>
<th>Table 1: Key Issues in Characterizing Uncertainty and Variability in Physiologically based Pharmacokinetic (PBPK) Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model specification:</td>
</tr>
<tr>
<td>- Integration of deterministic(^a) and nondeterministic(^b) model development</td>
</tr>
<tr>
<td>- Specification of alternative models</td>
</tr>
<tr>
<td>- Commonality of model structures across species</td>
</tr>
<tr>
<td>Model calibration:</td>
</tr>
<tr>
<td>- Use of data for estimating parameters versus “validating” the model</td>
</tr>
<tr>
<td>- Level of depth/rigor necessary in the nondeterministic model and parameter calibration methods</td>
</tr>
<tr>
<td>- Implementation of nondeterministic models (data inclusion/exclusion criteria, sources of variance/covariance to include, combined analysis of data with very different experimental designs)</td>
</tr>
<tr>
<td>- Evaluation of alternative models</td>
</tr>
<tr>
<td>Model prediction:</td>
</tr>
<tr>
<td>- Changes to the models and parameters for risk assessment predictions</td>
</tr>
<tr>
<td>- Characterizing uncertainty from alternative models</td>
</tr>
<tr>
<td>- Providing feedback to data needs and experimental design</td>
</tr>
</tbody>
</table>

\(^a\)The “deterministic” model is the mathematical representation of the biological/chemical system (e.g., PBPK model and metabolic scheme).

\(^b\)The “nondeterministic” model is the mathematical/statistical representation of the uncertainty, variability, and covariance of the data and parameters of the deterministic model (e.g., statistical model for measurement errors and population variability).
application, and is the first step in the development of any model. Specification of “deterministic” PBPK models has a long history, and typically begins with relatively simple model structures that are subsequently refined to incorporate new biological understanding or when data indicate inconsistencies with model predictions. For example, model structures developed for respiratory uptake in rats of volatile organic compounds such as methylene chloride (Andersen et al., 1987) worked well for many similar compounds, but did not capture the kinetics of other compounds without changes to account for additional pharmacokinetically relevant processes, such as sticking to fur, degradation of metabolizing enzymes, or induction of binding proteins (Gargas et al., 1990, Leung et al., 1990). Models fit to data from human methylene chloride exposure studies on the variability in respiratory uptake with workload (Johanson and Naslund, 1988; Jonsson et al., 2001a) were inadequate to predict acetone uptake because the model structure did not represent the “wash-in wash-out” behavior arising from distribution of this water soluble compound to the mucus layer in the respiratory tract (Mork and Johanson, 2006). Thus, discrepancies between data and model predictions highlight where better biological understanding is needed to motivate changes to the PBPK model structure to improve model performance. The modified structure then can be further tested experimentally.

The specification of “nondeterministic” (typically statistical) models for characterizing uncertainty and variability is standard practice in drug development using empirical PK models (e.g., one or two compartment models), but has not been routinely integrated into PBPK model development. Rather, statistical analyses of PBPK models are usually conducted post hoc. Nonetheless, there are a number of options for the nondeterministic/statistical component of a PBPK model, including those based on probability and likelihood theory (Knobloch et al., 2006), possibility theory, or fuzzy logic (Gueorguieva et al., 2004). Importantly, even simple least-squares optimization and “visual” fitting constitutes a “nondeterministic” model, although the latter is qualitative rather than quantitative.

Model Calibration

Once model structures are specified, values need to be obtained for the parameters. For purposes of this discussion, “model parameters” consist of those terms in PBPK models that generally are constant or exposure-independent during a simulation. Parameters, in this broad usage of the word, contrast with the changing chemical concentrations in tissues and excreta, which are calculated variables. In the deterministic part of the model, parameters include physiological parameters such as tissue volumes or blood flows and chemical-specific properties such as molecular weights or partition coefficients. In the nondeterministic model, parameters may include the population means and variances/covariances describing how physiological or chemical-specific parameter values differ among individuals to produce population variability. Other parameters may be used to reflect uncertainty arising from measurement errors or uncertainty in our ability to estimate the population variability. This is one of the many areas of potential terminological confusion for the disciplines interacting on this problem. Here a physiological parameter, e.g., lung volume, may possess a distribution in a population of interest. This distribution is characterized by one or more statistical parameters, e.g., mean and variance. The first use of “parameter” will be familiar to the PBPK modelers while the second use is more familiar to statisticians. The process of obtaining values for model parameters that allow the model to reasonably approximate the data is loosely termed “calibration.”

There have been several approaches to calibrating PBPK models. In some cases, there is a reliance solely on literature values for physiological parameters and in vitro estimates of partition coefficients and metabolic parameters. In other cases, varying levels of in vivo pharmacokinetic data have been used. To fit these data, “optimization” methods can be used, which generally refers to iterative processes of adjusting (typically a selected few) parameter values to minimize some function that quantifies the “distance” between model predictions and data (e.g., the sum of the squared residuals of the data from the model). “Parameter estimation” refers to statistically derived methods for calibration, such as maximum likelihood (Kalbfleisch, 1985) or Bayesian methods to estimate posterior distributions (Gelman et al., 1996), although in practice these methods typically rely on numerical optimization or sampling. While statistical estimation has many desirable properties, such as the ability to quantify the uncertainty and variability in parameter estimates, the approach is often complex and unfamiliar to many PBPK modelers. And, unlike the available statistical models (and software) for population PK modeling of clinical data, “standard” statistical models that have undergone full theoretical characterization do not (and are not likely to) exist for PBPK models. This difference reflects, in part, that PBPK models are high dimensional models with large numbers of parameters, while the empirical PK models used for population modeling have comparatively few parameters. The extent and kinds of data available for pharmaceutical versus environmental compounds also tend to be very different. Finally, with rare exception, current software packages emphasize either deterministic or statistical modeling, but not both.

Confidence in a model stems from both the structure of the model—the extent to which it reflects biological knowledge—and its ability to reproduce experimental data (Rescigno and Beck, 1987). In a research context, discrepancies can be useful in generating experimentally testable hypotheses as to additional biological processes that may be at work (Andersen et al., 2001; Kohn and Melnick, 1996; Leung et al., 1990). In risk assessment, where additional experimentation may not be possible in a timely manner, evaluating a model for consistency with existing data and acceptability for use presents many
challenges. (While the term “validation” usually refers to activities that build confidence that a model is appropriate for a given use, many workshop participants expressed discomfort at the term [but not the concept], as “validation” is sometimes taken to imply an attempt to determine the absolute “truth” of a model or is misinterpreted to imply that the model is appropriate for use in any application. For this reason, many preferred the word “evaluation” instead of “validation.”) No standard process exists for characterizing confidence in PBPK models, particularly with respect to comparison with experimental data. For instance, whether or not available data should be separated into “calibration” and “validation” data sets may depend on the availability of sufficient data and the calibration approach chosen. In principle, statistical tools (e.g., measures of goodness-of-fit, cross-validation) exist to determine if a model adequately describes data sets, while taking into account that parameters may have been estimated from those data, but there are few examples of such applications.

Model Prediction

In part due to differing needs between research (e.g., testing biological hypotheses) and risk assessment (e.g., estimating dose metrics used in quantifying risk estimates), the “handoff” of models from research to risk assessment applications has not always been smooth. Model components may need modification (e.g., different target tissues, interindividual variances), and one needs to ensure that equivalent predictions can be obtained for the data originally used to calibrate the model. Furthermore, a model that fails to adequately reproduce some data sets may yet be useful for some risk assessment purposes. A poorly fit data set may include quantities of little relevance to the specific exposure pathway of interest; for example, assessment of risks of oral exposure supported by data from oral dosing is not necessarily invalidated due to poorly predicted tissue concentrations following dermal exposure, though this limitation of the model should be noted. PBPK models are sometimes used to make predictions sensitive to parameters for which available experimental data do not provide direct information (e.g., estimating metabolic rates for a poorly metabolized compound from parent chemical plasma concentrations), so estimating the uncertainty in model predictions is important. It also may be useful to compare calibration results with predictions for a surrogate dose metric (e.g., area under the blood concentration curve) that may not be as closely related to toxicity, but that is more directly informed by experimental data (e.g., blood concentration), and, therefore, can be predicted with less uncertainty. Furthermore, even highly uncertain PBPK model predictions may be useful and informative as comparison values to “default” approaches (such as allometric scaling or uncertainty factors), which, while useful when looking across chemicals or in the absence of chemical-specific data, by their nature provide little information on uncertainty and variability for a specific chemical.

Given the diversity of approaches for specifying nondeterministic models, estimating parameters, and evaluating model performance, it is not surprising that multiple approaches have been used to help characterize uncertainty and variability in PBPK model-based predictions. As a preliminary step, sensitivity analyses can identify model parameters that are important for uncertainty/variability analysis. Point estimates can then be developed relatively quickly and may provide a convenient starting point when evaluating the impact of various assumptions on model predictions. Depending on the level of rigor required for the particular application, further analysis may not even be necessary.

Fuzzy numbers or other interval analyses are commonly used in engineering applications (Buckley, 1983; Singpurwalla and Booker, 2004) but are rarely used in PBPK applications (Gueorguieva et al., 2004). While probabilistic analyses are widely implemented (in particular via Monte-Carlo techniques), standardized methods have not been established to (1) choose distributions for parameters from published data; (2) set default distributions for parameters in the absence of data; and (3) account for correlations among parameters, though well-elaborated approaches are available (Clewell and Jarnot, 1994; Willmann et al., in press). Greater confidence in uncertainty and variability analyses will result if the parameter distributions/defaults, etc. are based on the best available biological understanding of the systems.

Finally, insights gained from model prediction should impact determinations of critical data needs and experimental design. Current practices have been relatively informal; using local sensitivity analysis to identify parameters of interest and relatively intuitive feedback to data needs. Optimal design, for which there are only a few examples (Bois et al., 1999; Brochot et al., in press; Müller, 1999), would incorporate the subsequent step of identifying specific data and experimental design characteristics that will efficiently identify the most sensitive parameters.

IMPROVING CURRENT PRACTICE AND IMPLEMENTATION

Multidisciplinary Integration of Deterministic and Nondeterministic Modeling

It became apparent at the Workshop that to better integrate the development of the deterministic and nondeterministic models, a multidisciplinary team of biologists, modelers, statisticians, applied mathematicians, and risk assessment users needs to be involved in all aspects of model definition, development, and implementation. This exchange is most effective if conducted throughout the modeling and risk assessment process rather than in an “assembly line” where PBPK modelers pass their model to the uncertainty/variability workers who then hand off their work to the risk assessors.
Achieving common understanding across disciplines of the concepts and terminology (e.g., uncertainty, variability, stochastic, etc.) is critical. This challenge was evident at the workshop. Participants were able to communicate well across disciplines after differences in definitions and their contexts were generally understood, even if not completely resolved.

For deterministic models, the team would collectively consider alternative model structures (i.e., alternative biological hypotheses), the forms of data available (e.g., literature summaries or data on individuals for physiology and pharmacokinetics), and the overall purpose of the modeling (e.g., dose–response analyses or extrapolations requiring parent compound/metabolite descriptions, specific target tissues and species, or potential dose metrics based on hypothesized mode(s) of action). This group would also define the distributions for physiological traits in the population of interest (e.g., lognormal distribution of liver volumes) along with the uncertainty associated with these distributions. These determinations would then guide implementation of the nondeterministic/statistical model, such as specification of likelihood functions, parameter distributions, parameters to fix or estimate, and the interpretation of results. This interdisciplinary approach would also include explicit specification of the key decisions made during model development and implementation, including evaluation criteria and the methods used to determine acceptable qualitative (visual) or quantitative fits to the data that, in turn, determine the adequacy of the model or prompt further changes in model structure or parameter estimates.

Broader Use of Sensitivity Analyses

Another common theme was the underutilization of sensitivity analyses to assess the impact of various options on model function (e.g., fits to data) and the impact of uncertainty associated with parameter estimates on model predictions. Sensitivity analysis can be useful in reducing the number of parameters considered by evaluating which parameters substantially influence model outputs (e.g., predicted dose metrics) under the conditions in pharmacokinetic or toxicological studies and human exposures of interest in the risk assessment. It can guide decisions to lump or split groups of processes (Gueorguieva et al., 2006; Nestorov et al. 1998). Notably, sensitivity analyses can be done in the context of formal estimation (i.e., as part of statistically based calibration), allowing one to look at the effects of modifying structures or parameters on the statistical likelihood used to fit the data.

In current practice, sensitivity analyses are typically univariate and local, in that they individually vary parameters by a small amount from “baseline” values, and assess the impact of each one on PBPK predictions as a function of time or dose (Lehman and Stark, 1982; Willems et al., 2001). However, covariances or interactions between parameters cannot be assessed in such analyses. These and related inferences as to the entire parameter space require multivariate, global sensitivity analyses (Brochot et al., in press; Gueorguieva et al., 2006; Iman and Helton, 1988; Saltelli et al., 2000), which examine the impact of varying all parameters simultaneously throughout a range of values that might be expected. Increased use of sensitivity analysis, including global analyses, was a near-term recommendation to provide a better characterization of the potential influence of parameter uncertainty and variability on model outputs.

Transparency, Reproducibility, and Completeness of Reporting and Documentation

A third major recognition was of the importance of comprehensive study reports that include an objective, logical, and transparent discussion of the choices made, the results of the sensitivity and other analyses, and access to all of the supporting, discrepant, or excluded data. Greater completeness in reporting would improve current practices, including the ability to independently reproduce the model (Andersen et al., 1995).

There was general consensus that objective and transparent sets of criteria are needed to guide decisions on how to include/exclude data, incorporate multiple sets of data, or evaluate the quality of models (Clark et al., 2004). These criteria should be based on sound statistical and biological principles, and developed within the context of the model structure and its application. Statistical expertise is useful both at the design and analysis stage, since the consequence of inappropriately accounting for data structures may include biased estimates, inappropriate estimates of variance, and/or questionable predictions. A stepwise approach to decision making may include the use of established and rigorous quantitative analysis of high quality data, to more qualitative methods (expert elicitation, decision analysis) due to inadequate data and insufficient resources or time to fill data gaps. However, equally apparent during the discussion was that generally applicable criteria are unlikely, so that any modeling effort needs to provide clear documentation of all the steps and decisions involved.

Uncertainty in the PBPK and statistical model structures has, to date, only rarely been addressed formally. However, even when only one PBPK model is “available” for a specific compound, it is usually the case that during model development, a number of other (typically unreported) alternative models were tried and failed to adequately capture aspects of the data (Willems et al., 2001). In most cases, these “discarded” models are not described. Thus, more complete documentation of the model development process would help address concerns about inadequate consideration of alternative models.

A major challenge to improved transparency is the lack of repositories for data, models, and their detailed documentation. Journals and editors could assist by providing places to deposit data for experimental studies, such as pharmacokinetic studies
on which modelers often rely, and supplementary model documentation, and by developing recommendations (or requirements) as to what should be archived.

**FUTURE DIRECTIONS: MEDIUM- AND LONG-TERM RESEARCH, DATA, AND IMPLEMENTATION NEEDS**

**Better Statistical Models and Methods**

Substantial progress has been made in applying statistical analyses to PBPK models, as evidenced by the increasing number of published analyses. A number of important issues in the characterization of uncertainty and variability in PBPK models remain unresolved:

- There is a need for better statistical models that both account for interindividual variability and the constraints imposed by laboratory animal studies (e.g., measurements at only a single time for any given animal, serial correlations).
- Consensus needs to be achieved on best practices for implementation of statistical modeling for PBPK models based upon data from multiple laboratories, often reported in aggregated form.
- There needs to be an exploration of the consequences with respect to model evaluation, parameter estimation, and model predictions of mis-specifying the statistical model or using less than optimal methods for estimating parameters.
- A parallel model calibration exercise with several groups estimating parameters for and evaluating the same set of models using the same data, but different methods, would be informative.
- More systematic methods for evaluating and comparing different models need to be developed, particularly for exploring the relative importance of uncertainties in model structure, dose metrics, model parameters, etc. along with different sources of variability.
- A more systematic understanding is needed of the impacts on uncertainty and variability estimates when maintaining a deterministic model structure across species with varying extents of experimental data in contrast to using reduced model forms for species where there are fewer relevant data sets for pharmacokinetic analyses.
- Metrics for characterizing the desired performance of PBPK models need to be developed.

**Better Databases for Physiological Properties and their Variation**

Workshop participants recognized a critical need for collaborations to improve peer-reviewed databases of parameters and covariates (such as lean body mass), particularly estimates and distributions for intra- and interindividual variability/covariance in both human and animal populations, prioritized according to the value of the information to meet specified applications.

**Explore a Wider Range of “Chemical Space”**

To date, the vast majority of PBPK models in toxicology have been for volatile organic chemicals using similar model structures. PBPK models are needed for a broader range of chemical space (e.g., from low to high volatility, low to high octanol:water partition coefficient). Chemicals with unusual properties can serve as probes of physiology (Andersen et al., 2001; Jonsson et al., 2001b), with the awareness that chemical exposures may also alter physiology. Methods to rapidly compare alternative model structures with data would facilitate testing of new structural ideas, provide perspective on model uncertainty, and help address “limited” data chemicals. Ultimately, the recognition that models of varying degrees of complexity may all describe the available data reasonably will encourage the acquisition of data to differentiate between competing models.

**Training, Documentation, and Software**

Finally, training materials and software improvements would lead to improved modeling practice and broader application to risk assessment. These include:

- Case studies of “proper” analyses and best practices
- Training for modelers on methods for statistical and other mathematical analyses (short courses, web-based tutorials)
- Bibliography compiled by topics (e.g., local and global sensitivity analysis, identifiability, lumping, model comparisons) with review and annotation to indicate strengths and limitations of published methods
- Glossary of terms including multiple, not necessarily consistent, uses of terms such as “parameter” as defined by statisticians and biological modelers
- Web-based model documentation repository.

**CONCLUSION**

Great strides have been made in deterministic PBPK modeling to characterize chemical pharmacokinetics and their implications for risk or safety assessment. Important challenges remain, particularly in their integration with nondeterministic/statistical models for analyzing pharmacokinetic uncertainty and variability. However, it is our hope that the multidisciplinary dialogue initiated by this Workshop will foster the collaboration, research, data collection, and training necessary to make characterizing uncertainty and variability a standard practice in PBPK modeling and risk assessment.

**ACKNOWLEDGMENTS**

The authors wish to thank the approximately 120 attendees for their active participation in the workshop. Participants who provided comments on
sections of this manuscript are too numerous to list by name. The authors thank Jerry Campbell for his assistance in manuscript preparation. This research was supported by the Intramural Research Program of the National Institutes of Health, NIEHS, the Office of Research and Development of the U.S. EPA, and the U.S. National Toxicology Program.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. EPA.

REFERENCES


