# An Umbrella Quality Assurance Project Plan (QAPP) for PBPK Models

Prepared

for

U.S. Environmental Protection Agency
Office of Research and Development
National Center for Environmental Assessment

Prepared by
PKWG QA Team

ORD QAPP ID Number:B-0030740-QP-1-0

## INTRODUCTION

The U.S. Environmental Protection (EPA) Agency requires project managers and planners to develop a Quality Assurance Project Plan (QAPP) as a tool for documenting the type and quality of data and model information that are needed for making environmental decisions. This document provides a QAPP that covers the basic data collection and modeling methodologies for physiologically-based pharmacokinetic (PBPK) models. It is an "Umbrella" QAPP and intended to be applicable to multiple PBPK modeling projects. This QAPP conforms to EPA QA/G-5 (U.S. EPA, 2002a) and is an internal QA Project Plan in support of the U.S. EPA's Human Health Risk Assessment (HHRA) research plan.

A PBPK model is a mathematical representation that describes the disposition of one or more chemicals in the body of a human or experimental animal in which organs or tissue groups are represented as compartments linked by blood flow that carries the chemical(s) between compartments. Put another way, a PBPK model is a quantitative statement of a set of hypotheses regarding the major determinants of absorption, distribution, metabolism and excretion (ADME). A key advantage of these models is that they can be used for various types of extrapolation including cross-species (animal to human), cross-route (e.g. inhalation to oral), and among exposure scenarios (Krishnan and Andersen, 1994), all of which can be used to facilitate human health risk evaluation and the setting of regulatory exposure levels. In addition to PBPK models, simpler pharmacokinetic (PK) with more empirically derived parameters can be used for the same types of extrapolation. Either form of PK model (PBPK models being a subset of all PK models) can be linked to a model describing some level of biological response, in which the combined dosimetry-response model is referred to as a biologically-based dose-response (BBDR) model. In this QAPP the term "PBPK" will be used since it is presently the most commonly evaluated and applied model form of model expected to be evaluated and used. However, this QAPP is intended to apply equally well to classical PK and BBDR model forms.

Guidance on the use or application of PBPK models in U.S. EPA risk assessments is not the subject of this document, but can be found in U.S. EPA (2006).

In order to adequately evaluate the quality of a PBPK model, a comprehensive understanding of a chemical's ADME processes (to the extent possible) is needed, which requires a review of any and all PK data available. For large PK data sets, an initial systematic review should allow for a selection of the publications which are most informative for PBPK modeling (i.e., a representative subset from among multiple publications containing similar information). For example, measurement of the total excretion in a 24-hour urine sample is less informative than time-course urinary excretion data. Further, if multiple publications report similar data, then the data from one or two of these might be adequate for model evaluation. The selection and grouping of representative data sets should be documented and checked under this QAPP, along with details of data extraction from the representative sets.

At the conclusion of the PK data review and QA, a set of key data files will then be available, where the methods and the data have been evaluated for accuracy and consistency with any apparent discrepancies resolved or explained. These should include any data sets used in development of the PBPK model, but may include other data. The model can then be evaluated against these quality-assured data, where the model parameters are expected to be consistent among the data sets, or to vary in a predictable way where data differ for explainable reasons.

Each model is defined by a specific set of model equations and a number of model parameters which must be chosen appropriately to match physiological, pharmacokinetic, or in vivo treatment-related response data. In order to be used with a reasonable level of confidence in human health risk assessment, the model must first be evaluated for quality, to assure that:

1) it properly represents the underlying biology, given the assumptions stated or implied in the scientific reference(s) describing the model (the model equations are correct);
2) the model parameters taken directly from the scientific literature have been transcribed accurately and appropriately applied;
3) the model otherwise fits or matches all of the QA data set (as described) with an acceptable degree of precision, or sound explanations exist for data that the model does not fit such that the discrepancy can be reasonably ignored.

Regarding (3), it is generally desirable that model predictions be within a factor of 2 of any data, but agreement of model predictions should be evaluated across an entire data set. Additional details on that aspect of evaluation are provided later in this document. An example where a large data discrepancy might be ignored is when the model describes some data quite well, in particular at exposure levels in the range of application, but does not fit other data (e.g., at higher, less relevant exposure levels). Alternately, the data may exhibit a high degree of variability that cannot be explained by strain, gender, or other experimental differences, making it impossible for any model to fit all the data with high precision.

It should also be noted that PBPK models include both code and associated parameter values and data. Because parameter values may be set or key calculations performed in in a model script separate from the file which defines the primary model equations, the term "model" as used here refers to the entire set or package of such files. An accurate model of human workplace exposure, for example, requires not only that the body weight, tissue fractions, and metabolic parameters be set properly for an adult human, but also the respiration rate and cardiac output expected in the workplace, and corresponding exposure levels as they vary during the day. The "model" is then the entirety of these equations and parameters, defined by all of the corresponding model files, which are addressed collectively by this QAPP

A model can be implemented in any of a number of software languages, such as ACSL (the primary language in which acslX models are written), R, MCSim, Python, Matlab, and Octave. A model implementation is the translation of the mathematical description, parameter values, and data used into one or more software languages, recorded in a set of computer files and scripts. These model files and scripts are then run or executed in a corresponding programming environment, which often is referred to by the same name as the language. For example, there is a Matlab language – the syntax and structure by which models are implemented in Matlab – and a Matlab environment where specific simulations and other model-based calculations are executed. This QAPP addresses the implementation of specific PBPK models and model applications into the corresponding computer files and scripts, irrespective of the language used.

Evaluation of the quality of the programming environments and evaluation and maintenance of their technical and user documentation is beyond the scope of this QAPP and is not the responsibility of the Pharmacokinetics Work Group (PKWG) (see A5: PKWG Background and Description) or individuals involved in developing and/or evaluating PBPK models as described here. Each of these programming environments is assumed to be fit for the purpose of scientific computing and documentation is assumed to be accurate as provided, though any evidence of errors or inaccuracy should be documented and reported to the software developer immediately.

On the other hand, if code packages (sets of files) or tools are developed to facilitate PBPK modeling for NCEA applications, i.e., that integrate with and extend a programming environment, then those packages or tools do fall under this QAPP, even if they are not PBPK models themselves.

For any specific PBPK project, an addendum to this umbrella QAPP may be produced that specifies additional details pursuant to its specific work plan.

## SECTION A: TASK MANAGEMENT

This section addresses task management including roles and responsibilities, background and description, quality objectives and criteria, training, documentation, and record keeping.

### *A1: Title and Approval Page*

Signatures indicate approval of this QAPP and a commitment to follow the applicable procedures noted therein.

V. Morozov 02/12/2018
Viktor Morozov, PKWG-Management Liaison — Date

Paul Schlosser Feb. 7, 2018
Paul Schlosser, PKWG Chair — Date

Yu-Sheng Lin Feb 12, 2018
Yu-Sheng Lin, PKWG member — Date

Feb 7, 2018
Dustin Kapraun, PKWG member — Date

2/12/18
Alan Sasso, PKWG member — Date

2/13/18
Cheryl Itkin, NCEA Director of Quality Assurance — Date

***A2: Table of Contents***

### *A3: Distribution List*

The individuals in Table 1, along with all members of the PKWG, will receive copies of the approved QAPP and subsequent revisions.

Table 1. QAPP distribution list.

| Name | Role | Organization | Contact Information |
|---|---|---|---|
| Paul Schlosser | PKWG Chair | ORD/NCEA/W | schlosser.paul@epa.gov<br>(919)-541-4130 |
| Viktor Morozov | PKWG-Management Liaison | ORD/NCEA/W | morozov,viktor@epa.gov<br>(703)-347-8156 |
| Cheryl Itkin | NCEA Director of QA | ORD/NCEA | itkin.cheryl@epa.gov<br>(703)-347-8557 |
| TBN | PKWG Co-Chair | | |

### *A4: Task Organization*

This section describes the roles and responsibilities for individuals associated with the PBPK Model QAPP.

#### **A4.1: Task Roles and Responsibilities**

The overall PBPK Modelling task includes the following roles: PKWG-Management Liaison, PKWG Chair (or Co-Chairs), NCEA Director of QA, and PKWG Project Leads (for specific chemicals or models; may also be the PKWG Chair); Principle Investigator (PI) or Contributing Investigator (CI) (for a specific sub-task; may be the Project Lead). The PKWG Project Lead is expected to manage the overall PBPK modelling task, including application of this QAPP.

Modelling tasks may also involve U.S. EPA staff or contractors not specifically identified in this document, but who are responsible for QA of the task or a sub-task; those individuals are referred to as the PI or a CI, depending on their role (for example, a contractor PI would be the primary or lead individual employed by a contractor as responsible for providing a work product), and shall be identified in an addendum, which can also describe other changes or additions to this QAPP for a specific model. While a PI or CI may in turn have assistance from other colleagues/staff, the QA responsibility shall not be delegated except to individuals identified as a PI or CI in an addendum. For example, a CI who is supporting the PKWG Project Lead may obtain assistance from a colleague (not identified in an addendum) in extracting data for a PBPK model, but that CI would still be responsible for documentation and QA of the extracted data.

*The PKWG-Management Liaison is responsible for the following:*

- approving the QAPP;
- providing an avenue of communication between NCEA management and the PKWG Chair(s);
- supporting the corresponding activities of the NCEA Director of QA and the PKWG Chair(s);
- ensuring implementation of QA corrective actions within the task when appropriate;

- facilitating project formulation, including defining desired outcomes and outputs; and
- working with NCEA management to allocate resources needed for model QA evaluations as described by this QAPP.

*The NCEA Director of QA is responsible for:*

- providing technical QA leadership for the task;
- ensuring all individuals developing or using the PBPK model have appropriate QA training (e.g., by taking the NCEA Quality Assurance Program Overview course in SkillPort);
- approving any supplemental model-specific amendments to the QAPP;
- advising the NCEA Director and other managers on QA-related issues requiring their attention;
- performing Technical System Audits (TSAs) ensuring corrective actions are completed as needed; and
- reviewing and approving QA documents generated by the task.

*The PKWG Chair(s) is (are) responsible for:*

- approving and updating the task QAPP;
- planning and identifying desired outcomes and outputs to be delivered for projects;
- ensuring the QAPP is implemented;
- ensuring quality related documents are developed and approved; and
- ensuring individuals in Table 1 receive the final QAPP and subsequent revisions.

*The PKWG Chair(s) and Project Lead are responsible for:*

- completing QA training;
- informing the appropriate chemical manager and others listed in Table 1 as appropriate of any model quality-related issues;
- obtaining the QA Manager's approval for quality-related documents;
- reviewing and approving the QA documentation (amendments) for specific models;
- participating in TSAs and implementing any corrective actions;
- ensuring that PIs or CIs are sufficiently qualified and that their contributions meet the QA objectives of this QAPP; and
- providing the model application and supporting documentation to the appropriate NCEA staff for subsequent distribution (e.g., inclusion of documentation in a Toxicological Review or posting of model code in HERO).

*PIs and CIs are responsible for:*

- ensuring that the QAPP is implemented as it relates to their specific tasks; and
- providing full documentation for their modeling tasks and QA activities to the Project Lead.

### *A5: PKWG Background and Description*

The PKWG was formed by the National Center for Environmental Assessment (NCEA) to support and promote consistent application of the best science practices in PK data analysis and modeling, including but not limited to PBPK modeling, as applied in human health risk assessment. The PKWG addresses the absorption, distribution, metabolism, and elimination (ADME) of chemicals in humans and laboratory animals, as well as the use or implementation of PBPK models. The objectives of this workgroup are to:

1. Promote and support the best use of available scientific PK data and methods in human health risk assessment in scientific products developed by NCEA for EPA's Integrated Risk Information System (IRIS);

2. As appropriate, promote and support the use of PK data and models in other scientific EPA assessments and products;

3. Advise NCEA management on issues related to PK data and modeling in human health risk assessment; and

4. Advance the scientific application of PK for human health risk assessment through further development and refinement of PK models, analysis methods and tools, and data resources.

A key part of the PKWG's work is to evaluate data covering ADME and PK, as well as the application of PBPK models for potential use in IRIS assessments and other EPA products. Occasionally new PBPK models are developed for use in IRIS, but it is more often the case that existing models are reviewed (including QA evaluation) and corrected or revised as deemed appropriate. According to the ORD Policies and Procedures Manual on quality assurance (Chapter 13, Section 13.9), this QAPP describes the QA documentation needed as a part of model development or review and revision.

### *A6: Quality Objectives and Criteria*

This QAPP seeks to ensure quality by establishing objectives and criteria for the development or elaboration; evaluation and correction; and application of PBPK models. The objectives for the QAPP include:

- providing a process that supports confidence and enhances transparency in scientific decisions based upon PBPK model application;
- creating a uniform framework for PBPK model development, revision, and QA review that is sufficiently flexible to encompass the various models, data sets, software environments, and the needs of the IRIS Program, NCEA, and EPA program and regional offices likely to occur;
- providing specific guidance for evaluation of PBPK models (e.g., selection of internal dose, incorporation of metabolic saturation as a function of dose or concentration); and

- increasing efficiency for risk assessment activities by minimizing the chances that errors occur in model code, parameter values, or data extraction and that where they exist (i.e., in existing models) they are identified and corrected quickly, as early as possible.

Because computational modeling requires specialized skills and knowledge, including familiarity with specific software, the conventions of PBPK modeling, and the related aspects of biology (physiology), biochemistry, and chemistry, the criteria which relate to clarity, transparency, or understandability of a model are understood to apply to an individual with a moderate level of PBPK modeling expertise. Quality criteria for this QAPP include:

- complete, transparent PBPK model descriptions (i.e., all equations, parameter derivations, and manipulations of data are completely and accurately described);
- model code and scripts that can be understood (i.e., have sufficient annotation) and used by an individual with moderate expertise in the selected programming language and environment;
- model accuracy and reliability:
    - equations in the model code accurately represent the model as described in supporting scientific papers, reports, or other documentation, with any discrepancies explained or resolved;
    - parameters in the model match those listed in supporting documents;
    - model parameters are accurately copied or extracted from scientific sources;
    - data used to calibrate or evaluate model predictions have been accurately copied or extracted from scientific sources; and
    - model results (numerical outputs, including tables and figures) can be replicated to at least the precision given in assessments and other documents where they are reported or used.

While not required, it is suggested that an accompanying 'readme' file be provided to guide model users and reviewers. The readme file should briefly describe each file in a code package; for example, the function or output of each accompanying script and the data contained in data files. For models in software environments such as R, where supplemental code packages in addition to the base installation are needed, and placement in the computer's file directory may be important, guidance on the installation should also be provided.

In addition to the quality objectives above, there are EPA, ORD, and NCEA policies and plans that guide quality activities including the following:

- ORD Quality Management Plan.
- NCEA Quality Management Plan (an appendix of the ORD Quality Management Plan).
- Chapter 13 of ORD's QA Policies and Procedures Manual.
- Requirements for Quality Assurance Project Plans (QA/G-5).
- Scientific Integrity Policy (www2.epa.gov/osa/policy-epa-scientific-integrity).
- EPA Information Quality Guidelines (www.epa.gov/quality/informationguidelines/).
- EPA Peer Review Handbook (www2.epa.gov/osa/peer-review-handbook-3rd-edition-2006-and-addendum).
- ORD Human Health Risk Assessment (HHRA) National Program Planning Documents.

## *A8: Documentation and Records*

This section discusses how and where the documents and records relating to a PBPK modelling task are maintained. The PKWG Chair(s), Project Leads, contributing EPA staff, and contractor PI should maintain documents and records associated with this task. For each PBPK model, or task associated with a PBPK model, a specific individual (either EPA employee or contractor) shall be identified, who will have primary responsibility for the documentation associated with that model or task. Documentation listed in A8.2: Documents and Records shall be maintained in EPA "cloud" storage, such as a OneDrive, SharePoint site, project site, in databases, or in version control repositories.

### A8.1: QAPP Distribution

The PKWG Chair(s) maintain(s) the final approved version of this QAPP in the NCEA-QA Team Documents share drive. The PKWG Chair(s) and PK Project Leads maintain a final version of any addenda for the chemical/assessment on which they are lead. The approved QAPP, including revisions, updates and any addenda are delivered electronically to the individuals listed in Table 1 and any other EPA staff of contractor PIs supporting PBPK modeling. The final approved QAPP and subsequent versions and addenda are stored by the PKWG Chair(s) or the PK Project Leads in the NCEA-ORD QA Team - Documents share drive.

### A8.2: Documents and Records

The PKWG Project Lead for a chemical/assessment, with assistance from any modeling PIs and CIs is expected to keep documents relating to the PBPK modeling task. Information includes:

- the source publication(s) or report(s) describing development of the PBPK model, in particular the choice of any features or model components that are not standard to PBPK modeling;
- the sources for all model parameters and data associated with the model, including page and table or figure numbers within a citation; these may also include spreadsheets or other files received from authors of publications and reports, but should generally be the sources cited in the paper or report where the PBPK model is described;
- comments or other documentation (e.g., spreadsheets or software scripts) sufficient to reproduce any conversion of published data to the actual values used in a PBPK model;
- the model code and any scripts (preferred) used to generate any plots, tables, or other results, or sufficiently detailed descriptions of the steps used to produce each plot, table or other result to allow it to be reproduced;
- a readme file (preferred) to guide model users regarding the primary components, features, and function of any scripts; the readme can presume a moderate level of PBPK modeling expertise;
- detailed descriptions of any changes made from a PBPK model as published in the scientific literature, including the rationale for the changes and indication of its impact on model predictions; and
- a master document that summarizes QA for the various pieces or individual files (e.g., QA checklists) in the set or package associated with each model.

This information should be maintained in shared electronic folders or databases. While copies of original publications or reports may just be kept in EPA's HERO database and only cited in the model's QA package, all other pieces should be organized together into a single folder or zip file at the end of the QA process.

### A8.3: Project Management Plan (PMP)

During the planning phase, the PKWG Project Lead works with project participants to create a PMP describing the work to be performed. Previous PMPs are maintained for audit purposes. All PMPs are provided to the PIs or CIs and saved as a part of the project's records.

### A8.4. Documents and Records Related to Peer Review

Typically, PBPK models are incorporated into IRIS Toxicological Reviews, in which they are peer reviewed along with the Toxicological Review as a whole. In this case, documents and records related to peer review are the responsibility of the chemical manager(s) and defined by that peer review process. If a PBPK model is being used as published in the peer-reviewed scientific literature, with only minor modification or corrections, then it is assumed that model was selected by the process described in the IRIS Handbook (i.e., by discussion and agreement among the PKWG, chemical managers, and other NCEA management personnel as appropriate), and no additional peer review (beyond that of the Toxicological Review) is necessary.

In the more unusual case that a model is being developed de novo, or a previously published model is being substantially altered by PKWG members or other NCEA staff (and submitted for journal publication separate from its use in a Toxicological Review), then an additional peer review process may be used. Corresponding documents and records may include:

- Internal Peer Review Plans (currently in Webforms, Form 117)
- Federal Register Notice(s), if generated
- Charge to Reviewers
- PBPK Model packages
- Logistical Fact Sheets
- Peer Review Reports or individual comments from reviewers.
- Disposition of comments.

These documents are kept by the PKWG Project Lead on his/her computer or the NCEA share drive and are the official peer review records. Public documents, such as manuscripts intended for peer-reviewed scientific journal publication, go through the NCEA clearance process using the STICS online database where copies of these documents can be found in that database.

### A8.5: Other Documentation

Other key documents associated with PBPK tasks include IRIS standard operating procedures, the IRIS Handbook, and software user manuals and documentation (i.e., provided by the software developer such as Matlab or R source-code providers). While it is not the responsibility of the PKWG, PI or CI to maintain master copies of these documents, it is helpful for them to have copies readily available on their computers or via internet links (e.g., to software documentation on the manufacturer's or developer's website).

## SECTION B: STANDARD OPERATING PROCEDURES (SOPs) FOR IDENTIFICATION, ORGANIZATION, AND EVALUATION OF ADME AND PK STUDIES AND MODELS

### *B1: DATA Review, Verification, Validation and Usability*

#### B1.1 ADME Data Evaluation and Selection

This section describes the analytical process by which information from ADME studies is evaluated and selected for use in PK modeling. Uncertainty in PK modeling is reduced when the most relevant, reliable, and quantitatively-valuable ADME studies are identified and given precedence over studies that provide limited information. It is important to identify all relevant, scientifically sound ADME data to provide the best possible basis for model calibration and evaluation. In particular, one would want to know how well a model describes any existing data, and the more data used in model evaluation and calibration, the lower the uncertainty in model predictions. On the other hand, for chemicals with very large available databases (there are hundreds of ADME studies for some chemicals), one will wish to identify a smaller, manageable set of PK studies that is representative of the larger database.

PBPK models serve to quantify inter- and intra-species PK differences, so are developed for specific animal species or humans. Therefore, the most relevant ADME studies are ones conducted in those species and it is generally acceptable to ignore studies from other species not being modeled. However, mechanistic information may be derived from other species, so a qualitative summary of those data can be helpful. ADME studies are used to: identify parent chemical and metabolite(s) found in test species and humans; demonstrate metabolic pathways; identify metabolizing enzymes and kinetic constants (e.g., $K_m$, $V_{max}$); characterize metabolic competition (i.e., when multiple chemicals compete for the same metabolic enzyme); characterize primary routes/methods of elimination; and identify data gaps toward which future research may be targeted. Given that nearly all PK reports have some level of intrinsic value, the considerations described below will help determine the level of detail at which these reports might be summarized.

For the purpose of PBPK modeling, optimal ADME studies are those that have been peer reviewed, have been conducted in humans or in the species/strain of animal being modeled, and have employed a range of doses that span those used in key toxicological studies or are relevant to human exposures. The most useful ADME studies report the time course for amounts or concentrations of a parent compound of interest and specifically-identified metabolite(s), providing information on the overall fate and mass balance of the parent chemical. For human ADME studies, doses in the range of the point of departure (POD) are ideal for informing animal-to-human extrapolation. Other studies, including those that evaluate formation of a given metabolite by in vitro methods may also have value.

While there is no formally established approach to categorize ADME studies based on their data type and depth of detail, a conceptualized "tiered approach" may be a useful tool through which to consider the value of each study. For example, the initial evaluation may focus only on the primary features of a study such as the species, strain, sex, developmental stage, exposure route and regimen of administration, sample timing, extent to which metabolites are identified and distinguished analytically from the parent chemical, and the number of time-points evaluated. The most promising studies identified by applying filters to this first tier of information (e.g., those conducted in the species, sex, strain, and developmental stage being modeled, and which are dosed via the route(s) of interest) can then be evaluated more carefully in a second-tier review for aspects of study and data quality. The second tier review might identify the studies which quantify levels of the parent compound and key metabolites, demonstrate the relationship between exposure and internal dose, provide time-course data in target tissues or blood, and employ sound analytical and statistical methods. The points

identified under the general considerations for in vitro and in vivo studies below should be used when evaluating study quality, whether or not a tiered approach is used.

It should be recognized that many chemicals produce multiple toxicities, through different MOAs, with different dose-response functions, and that a PBPK model may be used to help interpret results for multiple endpoints. It is recommended that ADME study and data selection focus not only on the apparent key effect (i.e., based on external dose-response and severity considerations), but other endpoints that are triggered by exposures within an order of magnitude of the most sensitive one.

The extent to which ADME reports address the following questions impacts their value for PK modeling. While answers to all these questions are not strictly required, they are all valid and useful for ranking such studies. For chemicals with many ADME publications, greater application of these questions will aid in selecting the best data for modeling.

*General considerations:*

- Have toxicity studies identified a responsive test animal species (e.g., Sprague-Dawley rat) and target organ or tissue (liver, thymus, kidney, brain)? Does the ADME investigation evaluate (tissues or samples from) the identified test species/strain or human? If not, to what extent can the species and tissue investigated be deemed an appropriate surrogate?
- Are the results based on chemical-specific identification and quantitation (e.g., gas chromatographic, high-performance liquid chromatography [HPLC], or mass spectral identification) or on general measures of chemical distribution (e.g., radiolabel quantitation)?
- For data from/in humans, is the characterization of exposure sufficient to inform qualitative or quantitative conclusions?
- To what extent can adverse outcome(s) be attributable to the parent chemical, metabolism of the parent chemical (via a specific pathway), or an identified metabolite? If the parent chemical or a key metabolite or pathway has been identified, to what extent does an ADME study inform the dosimetry of the parent chemical, specific metabolic pathway, or identified metabolite?
- To what extent can human data be used to characterize inter-individual PK variability?
- Are valid analytical methods utilized and described in sufficient detail to enable interpretation of the data; are limits of detection and/or quantification provided?
- To what extent has the report been subjected to a peer review? Is the document accessible in whole or in part?

*For in vitro ADME investigations:*

- To what extent has the concentration of the agent been localized (e.g., measurement in cells versus media) and characterized (e.g., parent chemical disappearance, metabolite formation)?
- Are non-biological sources of loss accounted for (e.g., volatilization, solubility, binding to non-biological test system components)?
- To what extent does the range of concentrations studied enable an evaluation of events at non-saturating and saturating conditions of metabolism, binding, or transport?
- What evidence is available to determine whether in vitro concentrations have in vivo relevance, both in studies conducted in animal models and in human environmental exposures?

- What is the biological level of organization of the in vitro system? How much extrapolation is required to convert from units observed (e.g., pmol product formed per minute per pmol enzyme) to values representative of the intact system? Do multiple bioprocessing steps or bifurcations in downstream or upstream metabolic process complicate the extrapolation?
- If metabolic rates have been determined using recombinantly expressed enzymes, has a relative activity factor been determined?
- If metabolic rate constants have been derived and presented by the authors, are the underlying data available for evaluation?

*For in vivo ADME studies:*

- Was the route and method (e.g., inhalation, oral drinking water, oral bolus) of administration consistent with the route and method of exposure used in the toxicity evaluations?
- How likely is it that differences between the vehicle used in the toxicity study and the ADME study may have introduced PK differences between the two studies?
- Is it likely that manipulations of the animal have altered the underlying anatomy, physiology, or biochemistry related to related ADME processes (e.g., could anesthesia have altered important functions like respiration and chemical metabolism)?
- Are time-course and/or exposure-dose PK data reported?
- What is the relationship of doses evaluated to the POD?
- Do the data demonstrate mass-balance? Or, do they focus on a single pathway or step in a complex overall metabolic pathway?

After considering the set of available ADME studies against the various factors described above, it should be possible to sort the studies according to their relevance to the intended PBPK model development and application (e.g., test species, route of exposure), type of information (studies that identify ADME mechanisms vs. those providing quantitative data useful for calibration and validation), and study quality. (which may enable ranking and selection of studies with apparently discordant results, or identification of those most useful for PK modeling).

In cases of apparently conflicting PK data sets, an analysis of the methods and details will be conducted to either resolve the discrepancy or decide which of the data sets is/are most likely to be correct. For example, there are sometimes significant strain- or gender-related differences in PK among laboratory animals. If apparent data discrepancies appear to be due to such differences, then a PBPK model would only be expected to fit a particular strain (or sex), and, for risk assessment application, this should be the one with critical dose-response data. Alternatively, model parameters might be identified for each strain, gender, life-stage, or other sub-population for which analysis is to be conducted. Discrepancies between data sets might also occur due to different analytical methods, in which case evaluation of the methods might lead to identification of certain data sets as unreliable. In each case, the rationale for selection or grouping of particular data sets will be recorded.

Once this is complete the qualitative information can be summarized (or used to evaluate the quality and completeness of an existing summary) and the studies from which data should be extracted for model calibration or validation identified. While it is beyond the scope of this QAPP to specify in detail how the summarization and study selection should be conducted, a written summary describing the approach used (e.g., tiered evaluation, with selection process at each tier) and the rationale for study

selection should be prepared, allowing for the process to be independently reviewed and possibly reproduced.

**B1.2: Extraction of Quantitative ADME Data and PK Model Parameters**

All sources of data and parameters used for model calibration and evaluation will be documented in text tables and/or Excel workbooks, with a level of detail to allow easy validation. In particular, specific table numbers, figure numbers, or page and paragraph/line numbers should be provided. If multiple entries in a table report alternate values of a quantity (e.g., measured by different techniques), then further detail shall be provided. If a model is obtained without documentation of ADME data and model parameters as described here, then such documentation shall be generated as part of the model QA evaluation.

*Model Parameters:*

Identifying the source of a PBPK model parameter as a publication describing a previous PBPK model where the parameter is in turn taken from an earlier source, is not sufficient, since that practice can lead to propagation of errors. The parameter value should be tracked back to and checked against the publication in which it is first reported or measured. This can include, however, articles and reports which comprehensively review and report physiological parameters, such as Brown et al. (1997) and ILSI (1994). However, for such comprehensive reviews, different values for the same parameter may be reported in different tables, hence it is particularly important to identify the specific table (and column/row) from which the parameter is taken.

Where calculations are used to convert reported parameters or data to values/units consistent with a model, sufficient detail to replicate the calculations shall be provided. Preferably, calculations and conversions are set up in computational scripts or Excel spreadsheets using embedded formulas. For example, if a tissue mass fraction is calculated from a reported tissue weight (TW) and body weight (BW), then the TW and BW are entered into adjacent columns, exactly as reported in the reference, and the resulting fraction (TW/BW) is calculated in a third column (e.g., the entry is '= C1/B1'), rather than entered as a numerical value. Comment text (and column headers in spreadsheets) would identify the data source(s), as described above and provide details for more complex calculations.

When parameters are derived by more elaborate means, for example a regression analysis, details sufficient to replicate the result should be provided; this can be readily accomplished by embedding the analysis in a script. Simple regressions can also be performed directly in Excel plots, with the equations shown, allowing for easy validation. If a regression is performed by other means (e.g., using the Solver function in Excel), then a plot of the resulting curve can be generated along with the data for visual comparison, which makes it immediately evident when a significant numerical error has occurred.

*Data:*

When data are received directly from the author, a copy of the data file shall be saved with "as received" and the date received or saved in the file name. Subsequent manipulations of the data file shall be done using copies of this original file, with that dependence documented in the copies or an accompanying text file.

If original data files are not available from the data authors (often the case for older data) then they should be validated against the published sources, with documentation generated in the process. Data provided in numerical form from an intermediate source (e.g., a model author) can be plotted and compared to a published figure as described below to ensure accuracy.

*Validation*:

All data and parameter extraction should be validating by having an individual other than the person who performed this initial extraction check the values against the original sources. If data were initially extracted by the authors of a publication, then a single reviewer (other than those authors) can

perform the check. For data sets with less than 20 entries, all entries should be checked. For larger data sets a minimum of 20 entries or 20% of the entries should be checked, whichever is greater.

When data are digitized from a published figure, a preferred method of validation is to plot the data in Excel using identical axis types (e.g., linear vs log) and scales and a clear background for the plot. This generated plot can then be placed on top of a graphic image of the plot from the publication, stretched or compressed to give exact alignment of the axes, but smaller symbol sizes/alternate colors in the generated Excel plot. It can then be quickly seen that the reproduced plot points exactly match those in the digital image (to within a few percent precision). If the initial extractor creates such a plot, then a reviewer only needs to visually examine the plot and check that the data values in the spreadsheet cells used by the plot match the values in files read or otherwise used for the model – the reviewer does not need to re-create the plot to check its accuracy.

## *B2: Review, Verification, and Validation of Existing Computational PBPK/PK Models*

### B2.1: General Approach for Model Evaluation

Criteria for judging the quality of a model provided here are separated into two categories: scientific and technical, which are respectively described in "B2.2: PBPK/PK Model Structure and Documentation (Criteria A)" and in "B2.3: PBPK/PK Model In-Depth Technical Evaluation (Criteria B)." In summary, the scientific criteria (primarily included in Criteria A) focus on whether or not the biology, chemistry, and other information available for chemical MOA(s) (or the subset of those being described by a specific model) are appropriately represented by the model structure and equations. The scientific criteria can be judged based on the (draft) publication or report that describes the model and do not require evaluation of the computer code. Criteria A also include preliminary technical criteria, such as availability of the computer code (if obtained from an outside source) and apparent completeness of parameter listing and documentation. The in-depth technical and remaining scientific criteria (Criteria B) focus on the accurate implementation of the conceptual model in the model code and scripts, use of correct or biologically consistent parameters in the model, and reproducibility of model results reported in journal publications and other documents. Any data sets incorporated into the model should be verified, and should be documented as described in B1.1 ADME Data Evaluation and Selection for their accuracy and quality.

While the criteria presented here are in part a component of the current IRIS process, similar scientific criteria have also been successfully applied and are described in greater detail by Chiu et al. (2007), McLanahan et al. (2012), IPCS (2010), and Clark et al. (2004). This approach stresses: (1) clarity in the documentation of model purpose, structure, and biological characterization; (2) validation of mathematical descriptions, parameter values, and computer implementation; and (3) evaluation of each plausible dose metric. Such transparency and documentation are important for compliance with the Agency's information quality guidelines (U.S. EPA, 2002b).

### B2.2: PBPK/PK Model Structure and Documentation (Criteria A)

It is assumed here that a journal article, report, or other scientific document describing the model structure, underlying science, and sources or methods for identifying all model parameters is available (need not be a peer-reviewed publication), and that a copy of the corresponding computer code has been obtained, along with permission for its use and subsequent public distribution. For QA evaluation, a brief report is prepared summarizing the key features of the PBPK model and its likely utility for use in a risk assessment. For example, one can quickly determine if a model has been calibrated for oral and/or inhalation exposures, and hence whether it is suitable for specific routes of exposure. This information is important for evaluating the potential applicability of a given PK or

PBPK model. For example, if it is thought that a key toxic endpoint results from metabolism to a reactive metabolite in a target tissue, then a model that doesn't predict that rate (dose metric) would not be useful. The model QA report should evaluate the following criteria, based on the model description in publications or reports.

*Scientific criteria for PBPK/PK models:*

- Biological basis for the model is accurate
  - Model equations are consistent with biochemical understanding and biological plausibility
  - Consistent with mechanisms that significantly impact dosimetry
  - Describes critical behavior, such as nonlinear kinetics in a relevant dose range
  - Predicts dose-metrics expected to be relevant and to be better correlated with toxicity or risk than applied doses
  - Applicable for relevant route(s) of exposure
- Model should describe existing PK data reasonably well
  - Shape: matches curvature or nonlinearity, inflection points, peak concentration time, etc.
  - Quantitative value: model predictions preferably within a factor of 2-3 of the data
- Validity of chemical-specific hypotheses:
  - Standard PBPK model compartments incorporate a limited number of hypotheses regarding ADME processes that have been tested and shown consistent with multiple data sets, for multiple chemicals, and therefore do not require in-depth consideration.
  - However, hypotheses specific to a particular chemical or chemical class, which are not supported by PBPK model agreement with data for other chemicals, should be evaluated more carefully, in particular when a hypothesis leads to prediction of much lower risk in humans than experimental animals.
    - For example, if it is hypothesized that a specific metabolic pathway operates in an experimental animal species (in a target tissue), making that species (tissue) particularly sensitive, then one should determine if there are ADME data for that metabolite (in the target tissue) in both sensitive and non-sensitive animal species demonstrating dosimetric differences commensurate with sensitivity, and dosimetric data in humans (or human tissues) demonstrating a lack of production.
    - Another example is the hypothesis that reactive metabolites formed in the liver will not have an impact on other tissues. But a moderately reactive metabolite with a half-life of minutes is sufficiently stable to be transported between tissues or cell types within a tissue, even if it is too reactive to measure in tissue samples from in vivo PK studies, so this hypothesis needs careful evaluation.
  - PBPK models which incorporate alternate hypotheses (e.g., some systemic distribution for a metabolite vs. none) may be equally consistent with the ADME data, but lead to very different risk predictions, and the resulting range of uncertainty should be considered.

*Technical criteria for PBPK/PK models (evaluate if scientific criteria are met):*

- Well-documented model code
- Parameters are clearly identified, including origin/derivation (validated as described in B1.2)
- Parameters do not vary unpredictably with dose (e.g., any dose-dependence in absorption constants is predictable across the dose ranges relevant for animal and human modeling)
- For probabilistic human models, evaluate parameter distributions in the model vs. full human variability. For example, Bayesian calibration applied to human data taken from only healthy adults, and with physiological parameters representing that group, may not be sufficient to describe the entire population. When specific factors such as a genetic polymorphism are known to impact human variability, an analyses which fails to incorporate them would not be considered sufficient to replace default uncertainty factors. Generally, all segments of the population should be included when evaluating the distribution of the Human Equivalent Dose (HED) or Human Equivalent Concentration (HEC), but limiting the analysis to only the most sensitive group can be considered.
- Sensitivity and uncertainty analysis has been conducted for relevant exposure levels (local sensitivity analysis is sufficient, although global sensitivity analysis is more informative)
    - If a sensitivity analysis was not conducted, then one should be performed as part of the QA evaluation
    - A sound explanation should be provided when sensitivity of the dose metric to model parameters differs from what is reasonably expected based on experience

**B2.3: PBPK/PK Model In-Depth Technical Evaluation (Criteria B)**

The following technical criteria address the computational implementation, including checking the code versus published or implied equations, and attempting to reproduce published figures and tables.

- Model equations and parameters specified in computer code match those published or implied[1] in the peer-reviewed manuscript or report
- Published figures and tables of model simulations are reproducible to within 10%
- The most rigorous approach to validating that a particular model implementation accurately represents the mathematical and conceptual model as described in a publication or report (or implied, if not all equations are explicitly listed) is to independently replicate coding of the model; e.g., in a different programming language/environment. Such re-coding, while not necessary for acceptance and application of a PBPK model, may also facilitate transparency and communication of the model for internal and external scientific reviewers and other stakeholders and interested parties.
- If errors in the model implementation (equations or parameters) are found and corrected, and the correction or change alters the evaluated model predictions (plots or tables showing model agreement with data) by less than 10%, the error is considered small enough to not invalidate

[1] Some publications assume familiarity with the standard forms or equations for PBPK model compartments and may only describe them in the text and provide the associated parameters, without listing the specific equations. In this case the equations are implied.

the model or any other parameter value, *even if model predictions outside the range of the data change by more than 10%*.

- Since model quality is judged by comparing model predictions to data, the impact of an error on model quality is evaluated only by determining the impact in the range of the data. The error is considered *de minimis*, hence acceptable, if the impact in the range of the data is less than 10%.
- An impact greater than 10% outside the range of any data may indicate uncertainty in model extrapolation to that range, but does not alter the evaluation of its technical quality.

- If scientifically justified, a new version of the model equation or parameter may be documented and used in place of a published version (even if errors/corrections in the original version do not result in changes greater than 10%)
- For corrections resulting in changes greater than 10% in the range of the data, or significant changes in model structure (vs. only revising parameters), the revised model should be evaluated as a new model version; key conclusions may be unchanged, but the quality cannot be judged based on results of the previous version.

**B2.4: Documentation of Model Evaluation**

Documentation of a model evaluation, in particular the technical evaluation (Criteria B) should be generated and saved on a network drive/folder specific to the model being evaluated, as described in section A8.2. A master check-list of items being evaluated (e.g., model parameters, model data, model equations) should be created, to include summaries of the initial evaluation, corrective actions, and final decision with respect to overall model quality or acceptability. For sets of model parameters or data, which can be large in themselves, dependent documents (checklists) can be generated. For example, the master check-list would identify "Model parameters" as one item, with a parameter check-list document identified therein. Evaluation of each parameter is then documented in the parameter check-list.

### *B3: Development of New PBPK Models, Significant Revisions of Existing Models, and Other Computational Analyses*

While section B2 specifically addresses the evaluation of existing PBPK models, development of new models, significant model revision, and other computational analysis (e.g., estimation of exposure from biomarker levels) should be subject to the same scientific criteria and conducted in a way that satisfies the quality criteria. Specifically:

- Parameters and data should be collected and documented consistent with section B1.2, with a second individual checking the values/extraction for accuracy.
- Complete details of unit conversions and other data manipulations, regressions, and the derivation of any non-typical model equations should be provided, with algebraic calculations embedded in Excel worksheets (using formulas) or in scripts (with comments).
- Model equations should be described in complete detail in a text document (e.g., a report or appendix), such that a reviewer can ascertain that the equations in the model code represent a correct mathematical translation of the model;

- comments should be provided within the code and scripts to facilitate review and QA (i.e., describing what lines or sections of code do) and at the top of model scripts to summarize their function;
- a second individual should check the model code and any accompanying scripts line-by-line to assure that the code matches the text description; or

- An accompanying "readme" file should be created to provide an overview and general directions for users. Instructions in this file should contain sufficient detail such that any person moderately experienced with programming and PBPK modeling can reproduce model results.
- Documentation of the QA evaluation in the form of tables or check-lists as described in section B2.4, listing all items checked, should be created and stored.

### *B4: Model Environment Conversion*

In order to support transparency and to facilitate external peer and stakeholder review of PBPK models, all such models should be made available in a freely available programming environment, such as R, MCSim, or Octave. If a model is already available in such an environment, then no conversion is required. However, when a model is converted from another environment it is expected that all numerical outputs (e.g., results reported in tables) and graphical outputs (plots) should be matched between versions. Numerical results should match to at least 3 significant figures and there should be essentially no observable discrepancy in graphical output, beyond those that result from formatting choices. In the process of checking and assuring this level of consistency between software environments, errors in model equations or parameters may be found. Thus, software environment conversion facilitates QA evaluation. Therefore, it may be desirable to convert a highly influential model to an alternate environment, or independently code the model in the same environment, even when that is not needed for model sharing and review. All files defining the model equations and parameters, and any other scripts for each equivalent model version, should be made available for review and evaluation.

## SECTION C: ASSESSMENT AND OVERSIGHT

This section describes quality assessments and other reviews that are conducted to determine whether this QAPP is being implemented as approved.

### *C1: Assessments and Oversight*

The PKWG is responsible for oversight for any ADME evaluation or PBPK modelling task being conducted in support of IRIS Toxicological Reviews. The PKWG provides overarching direction, ideas, and suggestions with respect to PBPK model-specific application features and methodology, although the primary work may be performed by other EPA modelers or contactors. The PKWG also evaluates PBPK model theory and the mathematical formulas used for model calculations, and reviews draft documents produced for ADME evaluation or PBPK modelling. With the agreement of NCEA management, the PKWG may also provide guidance, oversight, or direct support for PBPK modeling tasks being conducted by U.S. EPA program offices. The PKWG may also evaluate software platforms and provide feedback on usability, clarity, coding issues, and the correctness of application output, although full validation of large software packages is beyond the scope of this QAPP.

The NCEA Director of QA conducts TSAs on the PBPK task. The Director of QA may inspect electronic files and documents stored by the PKWG Chair(s) and Project Leads on their individual computers or shared network folders for the purpose of implementing this QAPP. Issues are discussed with the responsible individuals following the TSA. The PKWG Chair(s) and Project Leads, with assistance from any PIs or CIs, implement any corrections resulting from the TSA. The NCEA Director of QA monitors implementation.

### *C2: Reports to the PKWG and Management*

Copies of reports and draft documents evaluating ADME/PK or describing PBPK model development or revision, testing results, findings, and corrective actions developed to support IRIS Toxicological Reviews should be provided to the corresponding PKWG Project Lead and/or Chair(s). While the PKWG may not be providing direct support for a particular assessment, this communication will help the PKWG fulfill its oversight and review role, and to provide feedback on the materials in a timely manner.

The PKWG Chair(s) or Project Leads provide summary reports on QA reviews to the QRMG Branch Chief and chemical assessment managers as these are developed or completed. Meetings are held as needed, including other individuals working on the PBPK models and/or QA review, and others in NCEA management as appropriate, to discuss findings and how they will be addressed.

The NCEA Director of QA provides TSA reports to management. The TSA report includes areas of exceptional compliance and areas for improvement. The report also includes proposed corrective actions for findings included in the report. Any corrective action that is implemented and completed is reported to the chemical manager and others in management by the PKWG Project Lead or Chair(s).

### *C3: Federal Register Notices (FRNs)*

When draft or final PBPK models and supporting documentation materials are being announced in an FRN (i.e., for public comment), the PKWG Project Lead should check and assist with composing the draft FRN to assure accuracy. C4: Model Reconciliation with Needs and Intended Use

The PKWG Project Lead is responsible for identifying any aspects of a PBPK model that does not meet the objectives and criteria listed in this QAPP or the needs of the intended application. The potential strengths and limitations of the PBPK model should be communicated clearly. A discussion

with management may then occur to determine whether a model should be revised to address the shortcomings, or if the model application should be discontinued or adjusted to ameliorate the identified weaknesses or limitations.

## REFERENCES

Brown, RP; Delp, MD; Lindstedt, SL; Rhomberg, LR; Beliles, RP. (1997). Physiological parameter values for physiologically based pharmacokinetic models [Review]. Toxicol Ind Health 13: 407-484. http://dx.doi.org/10.1177/074823379701300401

Chiu, WA; Barton, HA; Dewoskin, RS; Schlosser, P; Thompson, CM; Sonawane, B; Lipscomb, JC; Krishnan, K. (2007). Evaluation of physiologically based pharmacokinetic models for use in risk assessment [Review]. J Appl Toxicol 27: 218 - 237. http://dx.doi.org/10.1002/jat.1225

Clark, LH; Setzer, RW; Barton, HA. (2004). Framework for evaluation of physiologically-based pharmacokinetic models for use in safety or risk assessment. Risk Anal 24: 1697-1717. http://dx.doi.org/10.1111/j.0272-4332.2004.00561.x

ILSI. (1994). Physiological parameter values for PBPK models. Washington, DC: U.S. Environmental Protection Agency.

IPCS (International Programme on Chemical Safety). (2010). Characterization and application of physiologically based pharmacokinetic models in risk assessment. (Harmonization Project Document No 9). Geneva, Switzerland: World Health Organization. http://www.inchem.org/documents/harmproj/harmproj/harmproj9.pdf

Krishnan, K; Andersen, ME. (1994). Physiologically based pharmacokinetic modeling in toxicology. In AW Hayes (Ed.), (3rd ed., pp. 149-188). New York: Raven Press, Ltd.

McLanahan, ED; El-Masri, HA; Sweeney, LM; Kopylev, LY; Clewell, HJ; Wambaugh, JF; Schlosser, PM. (2012). Physiologically based pharmacokinetic model use in risk assessment--Why being published is not enough. Toxicol Sci 126: 5-15. http://dx.doi.org/10.1093/toxsci/kfr295

U.S. EPA. (2002a). Guidance for quality assurance project plans. (240R02009). http://nepis.epa.gov/exe/ZyPURL.cgi?Dockey=20011HPE.txt

U.S. EPA (U.S. Environmental Protection Agency). (2002b). Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity, of information disseminated by the Environmental Protection Agency. (EPA/260/R-02/008). Washington, DC: U.S. Environmental Protection Agency, Office of Environmental Information. http://www.epa.gov/quality/informationguidelines/documents/EPA_InfoQualityGuidelines.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2006). Approaches for the application of physiologically based pharmacokinetic (PBPK) models and supporting data in risk assessment (Final Report) [EPA Report] (pp. 1-123). (EPA/600/R-05/043F). Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=157668