Simulation in Drug Development: Good Practices

Document created in 1999 and published on the CDDS web site.

Editors
Holford NHG, Hale M, Ko HC, Steimer J-L, Sheiner LB, Peck CC

Contributors
Bonate P, Gillespie WR, Ludden T, Rubin DB, Stanski D

1 INTRODUCTION

With the rapidly changing health care and research environments, it has become essential that the drug development process achieve greater efficiency, cost-effectiveness and timeliness. A major component of drug development is drug testing in human for safety and efficacy. Past approaches to clinical drug development often resulted in much of the human clinical trials information being considered less than maximally informative, providing results which did not add new information that was relevant to the drug development and the approval process (Peck 1997).

Simulation of clinical trials has recently gained attention as an emerging technique for knowledge synthesis and exploration of possible clinical trial results based upon a mathematical/stochastic model of the trial, including sub-models of the drug action and disease process (Hale et al. 1996, Peck & Desjardins 1996, ECPM 1996, CDDS 1997, FDA 1999, Peck 1997b, Krall et al. 1998). The basic rationale for computer simulation has existed for many years and the technique has been successfully used in several scientific and industrial application areas (Johnson 1998). The application of simulation in the domain of pharmaceutical medicine, clinical pharmacology and drug development has been largely restricted in the past to evaluation of statistical methodology and forecasting of individual or population pharmacokinetics. It is proposed that simulation has a much broader potential to aid in the clinical development, regulatory review, commercialization, and medical application process. As described in detail later, models will incorporate elements associated with the drug, the disease and the trial such as study design, dosage regimens, population pharmacokinetics and pharmacodynamics, disease progression, placebo response, compliance patterns, dropout rates, study end-points, sample schedules and statistical analysis approaches. The primary purpose of this new approach is to improve clinical development by generating better insights into the consequences of the choices made in the design of human trials, especially at the planning stage.

2 GUIDING PRINCIPLES

This document is motivated by the belief that a clear and public articulation of agreed upon “Good Practices” can aid in the development and application of model building linked to simulation of clinical trials. A shared language and defined approaches can provide a basis for meaningful communication among scientists and clinicians in the area of drug development. It is expected that the general acceptance of agreed upon good practices will advance the state of the art, promote better utilization of

the methodology and allow non-experts to understand better and usefully apply the results of a simulation investigation. The definition of “Good Practices” aims at the following principles:

CLARITY: The report of the simulation should be understandable in terms of scope and conclusions by intended users such as those responsible for committing resources to a clinical trial.
COMPLETENESS: The assumptions, methods and critical results should be described in sufficient detail to be reproduced by an independent team.
PARSIMONY: The complexity of the models and simulation procedures should be no more than necessary to meet the objectives of the simulation project. Program codes sufficient to generate models, simulate trials and perform replication and simulation project level analyses should be retained but there is no need to store simulated trial and analysis results which can be reproduced from these codes.

3 PLANNING A SIMULATION PROJECT

One of the first tasks in approaching a simulation project is to identify clearly the purposes of the activity and the consumers of the information provided by the project, typically the company-internal teams, but possibly also regulatory scientists who may be consulted about the trial. Brief projects that are mostly exploratory in nature, involving few consumers, may have very modest needs for a plan of work. Most projects will require significant effort, however, where a thorough plan of work is needed for communication, efficiency, coherence of approach and, last but not least, for ensuring that adequate time and resources (in both manpower and computing) will be allocated. Just as a builder should have blueprints before starting construction of a new building, those undertaking the task of representing a clinical trial in mathematical terms and software code would do well to use a plan defined with a level of rigor that permits peers to examine the assumptions and approach. Doing this provides some assurance that the specified needs will be met. In the remainder of this document, we refer to this as a “simulation plan”.

3.1 Simulation Team

The simulation objectives and aims should influence the composition of the simulation team. The core team would comprise a specialist clinician (who may also be a trial investigator), a clinical pharmacologist, and a statistician. In case the mathematical modeling expertise is not adequately covered by these individuals, the presence of a pharmacometrician appears mandatory, especially in those cases where the PK and PD aspects are a substantial component of the model. At least one of these scientists must, obviously, have the talent to convert ideas, assumptions, premises as well as

mathematical/statistical models into software code. Other expertise may be necessary. For example, if there is a goal related to expected health-care costs, then a team member qualified for econometric modeling is needed. As simulations grow more complex and encompass multiple objectives, the simulation team will grow to an even greater level of cross-functionality.

3.2 Simulation Plan

Just as the protocol for a clinical study describes objectives, hypotheses and assumptions, trial parameters, methods, and analyses, so should the plan for a simulation project. The aim should be to produce a written document, with enough detail that another researcher can obtain comparable (simulation) results by following the (simulation) plan. Care in preparation of the plan will provide the basis for critical review of the components of the simulation project, and will assist in implementation of the computer simulation.

Development of the plan before commencing the numerical simulations provides a good opportunity for critical evaluation of assumptions, methods, and goals by team members, and gives some protection against analysts personal biases or "unreported discarding of models that didn't work". The plan defines the path for the simulation project, and provides for pre-agreed criteria against which the simulation results will be assessed. This discipline is particularly useful for computer simulated trials, where the relatively low cost of additional runs can lead to unreported "tweaking" of assumptions and models, chasing results, and leading to self-deception.

3.3 Overall Objectives and Specific Aims

The explicit statement of overall objectives for the simulation project provides a basis for all decisions and actions related to the project. Objectives and specific aims should be clearly stated in the simulation plan and agreed upon by shareholders before the simulations are performed. The specific aims will determine the selection of models and methods, and their implementation. For example, if a primary objective is to estimate what proportion of patients may be expected to experience a certain adverse event, then the sample size and methods proposed for the simulation will need to be sufficient to estimate that proportion to a desired precision.

3.4 Assumptions

Assumptions comprise essentially all components of the simulation model. Examples include structure of the models for pharmacokinetics (dose-concentration), pharmacodynamics (exposure-effect), clinical

effect, and covariate influences, their parameter values, and attendant variance structures. Further examples are assumptions about deviations from prescribed (clinical) protocol, which capture features such as non-compliance with treatment and study dropouts and how they may impact the trial. The assumptions should be explicitly identified in the simulation plan. If some models are incomplete at the planning stage, that should be noted, with a plan for model completion and later plan revision. Although not the most desirable situation, this is necessary in complex situations where a sequential approach to simulation is needed.

It is important to acknowledge several levels of assumptions based upon level of underlying evidence or knowledge: 1) data & experiment based 2) educated or theoretically justified 3) necessary for the simulation but largely conjectural and may be the focus of the simulation experiment. This holds for all sub-models that are described in more detail in Section 4. Premises of lesser certainty should be considered for inclusion as factors (see Section 3.4) to be varied in the simulations. Premises of greater certainty might remain unchanged throughout the simulation, possibly stipulated as "true" or at least widely accepted as so.

3.5 Design of the Simulation Project

Clinical trial simulation will often be approached as an experiment (an "in silico" or “in numero” experiment), where factors are varied to determine their impact on outcomes. These factors include trial design properties (see Section 3.6.3), simulation models, and their parameters, (see Section 3.7). Factors may take on specified values, or the value taken may be sampled randomly from a probability distribution. “Fractional" or "response surface" designs (Box et al. 1978) are often a good choice since they provide an efficient and well understood way to examine relationships between many factors and outcomes.These designs may be used to provide maximum reliability from the amount of resources devoted to the project, and allow for examination of individual and joint impact of numerous factors, rather than relying on relatively inefficient “one-factor-at-a-time” experimentation.

The factors and their combinations should be identified for the experimental design. Factor ranges and probability distributions should be specified. Outcomes also need clear definition, usually at multiple levels, e.g., individual patient outcome, treatment group outcome, trial outcome. When the simulation is to represent a real trial, reference to the outcomes as defined in the real trial protocol is essential. If one purpose of the simulation is to help develop the real trial protocol, such as defining entry criteria, demographic characteristics, study variables of primary interest, times of observations, etc., then the possibilities under consideration are good candidates for investigation as factors in the simulation project. Each simulated trial should be replicated sufficient times to meet project objectives (see Section 3.6.2). For example, far fewer replications will be needed to evaluate median behavior than to evaluate tail

behavior (e.g., distribution of values for small percentiles). Estimates for a suitable number of replications (i.e. the “sample size” for a given simulation "experiment") will often be approximate because of the complexity of the simulations, but can be estimated more precisely from initial simulated results. Simulation provides for systematic evaluation of the properties of alternative clinical trial designs. Consideration should be given to whether study costs should be incorporated and tracked during the simulation, as this could well vary with design, and might be the deciding factor in designs that similarly on other counts.

3.6 Simulation Project Design

3.6.1 Experimental Design

The experimental design for a simulation, in many ways, is just like the experimental design for an actual experiment, with two primary differences: First, because of real resource limitations, some factors do not vary in the actual experiment (e.g., the number and type of subjects, Latin square vs. Greco-Latin square), whereas in a simulation experiment, such factors can be varied as part of the simulation project in order to investigate experimentally the effect of design properties on results (see Section 3.6.3). Second, nature generates the responses (or “outputs”, or “outcomes”) in an actual experiment, whereas the computer, through implementation of a simulation model, generates the responses in a simulation experiment.

The selection of the factors that the trial simulation team wishes to vary in an actual experiment, such as dose, is essentially the same task in simulation experiments and actual experiments. However, the selection of factors and their levels describing models used in generating data is a task for the designer of simulation experiments that is accomplished by nature in actual experiments. For this reason, the design of a simulation experiment involves more factors than the design of the corresponding natural experiment. Factors in simulation experiments for generating responses correspond to models for at least three distinct aspects of nature: first, input-output models describing how the outcomes vary as a function of the background variables and treatment exposures (e.g., PK/PD models); second, covariate distribution models describing the background/baseline characteristics of the population from which the simulated trial subjects will be sampled (e.g., age, sex, race, blood pressure, cholesterol concentration); third, execution models describing how deviations from protocol, such as noncompliance and missing data due to non-response, transform the nominal design of a trial (as planned in the clinical trial protocol) into the actual design (as arising from the actual conduct of the trial).

Sources of information for these three models are very different, but all are needed to realistically simulate how nature produces outcomes in an actual experiment. The IO model will rely on the usual set of pre-clinical and clinical scientific studies with the drug, and the body of literature (including published

models) on related compounds. The covariate distribution model primarily relies upon available population data bases. The execution model relies on information on actual behavior of individuals derived from experience with real world experiments.

The number of possible designs may be overwhelming. Accordingly, efficient means of exploration must be employed, thus introducing the requirement for a "design" of the simulation project *per se* or what might be called a *meta-design,* (which differs from the design of the clinical trial, which is the subject of the simulation project). In this situation, it is even more critical, than in an actual experiment, to capitalize on ideas from the statistical sub-field of experimental design with factorial experiments (Sacks et al. 1989a, 1989b, Welch et al. 1992) . In particular, *fractional* replication may be relevant to defining the meta-design, because its purpose is to create efficient designs when the presence of (too) many factors to investigate prevents the use of a complete factorial design (i.e., all combinations of factor levels being studied). Response surface designs may also be employed in designing simulations aimed at finding a nearly optimal actual experimental design, especially when a sequence of simulation experiments can be contemplated.

3.6.2 Replications

The number of replications (i.e., the number of simulations of an individual trial). should be justified by the objectives and precision required of the simulation. An estimate may be derived via formal (statistical) calculations as well as pilot simulations. If the variable(s) of interest is (are) discrete, the number of simulations can be calculated from all possible combinations of outcomes using combinatorial algebra to estimate the number of replications and time required. Further, when the end-result of a Monte Carlo simulation is the calculation of some p-value or the percent of simulations rejecting some null hypothesis, which are binomially distributed, then the variance of that statistic can be estimated as p(1-p)/n, where n is the number of replications. This equation can be rearranged and some pilot simulation data can be incorporated to calculate a specified degree of precision in the estimate (normally distributed). Further pilot simulations may lead to either expanded or reduced scope, as resources permit. For continuous variables, standard power calculations for a desired level of precision can be done. It is important to be able to output the results of each replication to a file, such as an ASCII file, to permit further analysis of the full set of replications.

3.6.3 Trial Design Properties

The controllable variables of the trial design may be referred to as design *properties.* This term usefully distinguishes them from other simulation variables such as the *parameters* of the various sub-models

used for the simulation. Design properties can be broadly related to the subject population, the treatments and the observations.

Population properties are used to select subjects from the population covariate model (see above); e.g., ranges of age, weight, renal function or the proportion of males and females. These properties are used to implement those design features usually described as inclusion and exclusion criteria in clinical trial protocols.

Treatment properties reflect the number of subjects assigned to each treatment group and the nature of the treatment for each group such as the dose size, formulation and dosing frequency. The kind of treatment assignment, e.g., parallel group, cross-over, forced titration, or dose-escalation, is also a treatment property that is often the most crucial feature of the overall design. The method of assignment is another treatment property, but this is almost always method of randomization.

Observation properties specify the type of responses (biomarker, surrogate or clinical endpoint) to be measured and the number and timing of each observations.

The selection of a set of trial design properties uniquely identifies a particular design. One replication of a particular design yields, after statistical analysis of raw trial results, a summary statistics (perhaps simply the p-value of a test of the null hypothesis). The performance of the design is sometimes judged in terms of the cumulative distribution of such a statistic, e.g., the probability of rejecting the null hypothesis under the alternative hypothesis; i.e., power. To find a good design (or, even more difficult, a robust design) the selected designs must be evaluated (see Section 3.6.1).

3.7 Models for Simulation

Constrained by the ***parsimony principle***, the type of models employed may have both empirical and mechanistic elements. Sub-models should be identified, with appropriate literature references when such exist. The models will typically have both fixed and random components. Multivariate distributions should be used when possible rather than independent univariate distributions, e.g. age and renal function, especially for characteristics in the targeted population that are highly correlated.

The simulation team must consider which models for dose-response or concentration-response (efficacy and safety), compliance, dropout, etc. would enable a realistic simulation. A "full blown" system model may not be needed to meet simulation objectives. Here, again, the parsimony principle should be considered so that over-complex models are not used.

Most modelers of biological phenomena are familiar with so called “Input-Output” (or “IO”) models, i.e. models that predict responses (or “outputs”, or “outcomes”) given certain inputs and baseline covariate values. In PK/PD, the outputs are drug concentrations and effects; the inputs are rates of drug input over time, and the baseline covariates are such things as species of animal, age, gender, values of laboratory

tests, etc. When the IO relationship involves stochastic elements (e.g. between and within subject variability and measurement error), a complete model must also describe the ***probability distribution of outputs*** given inputs. It is customary to think of the ***expected (mean) output*** of the IO relationship when the word "model" is employed. In the context of this document, when we use the unmodified term "input-output model," we refer to a full probability model; that is, a model for the entire probability density function for the outputs as a function of the inputs (or probability mass function for discrete outputs). Critical parameters that describe these models will be among the factors to be explored in the simulation project.

The sub-models required to simulate clinical trials include an IO model (see Section 3.7.1) and two additional models (covariate distribution model (Section 3.7.2) and execution model (Section 3.7.3)), which are often unfamiliar to modelers involved primarily with data analysis. A clinical trial can be thought of as a series of steps, each involving, for simulation purposes, a sub-model from which the outputs particular to that step must be generated. Those steps are 1) creation of a study population, 2) selection of study design, 3) trial conduct, and 4) analysis of trial results. Those steps are explained in more detail in the following paragraphs.

First, a study population is created. Simulated subjects must be drawn from a probability model for baseline covariates describing that of the intended (real) subject population. The probability model of population characteristics should include those characteristics that are known or suspected to be of relevance (age, race, weight, disease state, etc.). Anything described by entry criteria in the actual trial protocol falls into this category, as these are needed for determination whether these subjects are allowed into the simulated trial or not.

Once a probability model describing the population of subjects has been chosen, a nominal study design is selected (see Trial Design Properties (Section 3.6.3)), which fixes the value of the controllable design properties. The nominal design does not arise from a model, as, in general, it has no stochastic elements: it is determined by the choice of the study design team as to the settings of the design properties. (Note, however, that so-called "adaptive" designs, ones that change depending on observed outcomes, indeed do have stochastic elements, and therefore require a model in the sense of the word used here). It is often the primary purpose of a simulation study to inform the trial designers' (and also the experimenters') judgment in making those choices.

The next step is the trial execution. The Execution model will use the *nominal* design to simulate an *as executed* design which reflects events such as compliance variation and subject drop-outs. Given the *as executed* design (note, not the *nominal* design) and baseline covariates, the Input-Output Models for the outcomes provides the results of the simulated instance of the clinical trial. The IO model outputs are then analyzed according to the method specified for analysis of the nominal design for the individual simulated trial (this step just reflects the current practice of statistical analysis). This analysis is the one that is made

explicit in the Analyses section of the simulation plan (see section 3.9). The model for the transformation of the executed design to outcomes is the familiar IO model, often one linking drug dosage to clinical outcomes via a series of PK and PD sub-models. Application of this IO model to the process of clinical trial simulation differs from the application of such a model to data analysis in several ways.
For simulation purposes, the IO model will differ even from population models as currently used for analysis of actual trials, in the attention that must be paid to reflect faithfully the variability in the data. As stated elsewhere (section 3.9), models for simulation are more complex than for analysis. Population models are traditionally used to draw conclusions from already-completed trials about the sub-model for the expected value of individual IO model parameters as a function of covariates. For those purposes, while inference demands that the within-individual correlation of measurements be recognized, in general, the sensitivity of estimates to the accuracy of the sub-models for variability is not great. In contrast, a simulation project will often seek to estimate also the sensitivity of trial design performance with respect to tail probabilities of events, such as the distribution of responses at a given time after therapy begins, and so the joint-probability models giving rise to such events must, accordingly, be well represented.

3.7.1 Input-Output Models

IO models may be broadly divided into mechanistic and empirical models. Mechanistic models attempt to reflect, at a structural level, the actual physical/biological system giving rise to the data, whereas empirical models simply describe the shape of the IO relationship. For simulation studies, mechanistic models are encouraged. Such models are expected to extrapolate to new situations better than empirical models, and exploring the study design properties in a simulation project inevitably requires extrapolation beyond current data. Attention should be paid to exploring responses which may arise from abrupt withdrawal of drug or brief drug holidays. Such phenomena may however be difficult to simulate because of the relative paucity of plausible models.

3.7.2 Covariate Distribution Models

At a first level, covariate models define the distribution of covariates in the population to be studied in the trial. The relevance of these is that IO models used for simulation studies must deal with the variability from individual to individual, and within individuals over time. Models that can do this must account for a rich and complex co-variation between observations within individuals. For mechanistic models, such complex modeling of co-variation is best done using so-called hierarchical random effects models, which view the parameters of the individual-level IO models as themselves random, with distributions governed

by baseline covariates (hence the need for the covariate model) and, perhaps, certain outcomes (including measurements of the same covariates observed at baseline). In a sense, an IO model with this feature is also a model of the population: it accounts for the distribution in the population of parameters governing the individual IO models, often as a function of baseline covariates. Such models have become familiar to PK/PD researchers as so-called “population models”, and to Bayesian statistical data analysts as “hierarchical models”.

3.7.3 Execution Models

During execution of the real clinical experiment, deviations from the trial protocol will inevitably occur. To simulate a clinical trial with proper accounting for such perturbations, a model of deviations from per-protocol behavior must be put forward. These will consist of individuals who refuse to enter the study or are inappropriately included or excluded (initiation deviations), those who do not comply fully with instructions (they miss doses, clinic visits, etc; so-called compliance deviations), and individuals who drop out of the study prematurely (termination deviations) (Urquhart 1999) . Deviations may also be attributable to investigator behavior such as failing to obtain an observation or not recording the time of the observation accurately (observation deviations). Such models are often unfamiliar to modelers, primarily because laboratory scientists usually deal with experiments in which deviations are minor or absent, whereas those who analyze clinical trials, usually use the standard approach to such analysis (i.e., “intention to treat”), which ignores such deviations. To simulate a clinical trial realistically, however, the data must be generated from a realistic simulation model, no matter what method of analysis is applied. The inputs to a model for deviations from protocol are the nominal design, baseline covariates, and outcomes. Of course, account must be taken that only those outcomes that have occurred before the time of a given protocol event (that may or may not be executed properly), can influence the resulting event.

3.7.4 Source of Models

Ideally, the models needed to perform simulation studies of the next series of actual clinical trials to be undertaken are developed during the course of prior investigation with the drug. Thus models for phase 2 drug development trials should be developed in phase 1, for phase 3 in phase 2, etc. This flow of development means that one criterion for a trial design, arguably the most important at any phase, is the ability of that design to reveal the models (required to simulate the next stage) with sufficient accuracy and precision for reliable decision-making. Of course such models will inevitably remain uncertain; the effect of this uncertainty on design performance is the subject of sensitivity analyses: one seeks “robust”

designs; that is, those that will perform well under a variety of premises, as translated into the simulation via model uncertainty. Although information from previous development phases may serve well to express the drug-specific IO models, it will not, in general, be adequate to provide information on the covariate distribution model or the execution model for deviations from the clinical trial protocol.
Covariate models will be largely empirical, not mechanistic, and should be constructed and estimated from existing databases of covariate values in populations of interest. A problem with this approach is not only the availability of data bases for public use, but the incompleteness of any particular data base: not all covariates of interest are measured in every study. Modern methods of data imputation, adjusted so as to allow correction for the degree of imputation in subsequent inference are available, and may find application here (Rubin 1996). Such databases need not, of course, come solely from therapeutic studies: data bases from health care systems, for example, should be quite useful here (e.g., National Health and Nutrition Examination Survey).
Models for deviation from clinical trial protocol will be more difficult to specify with any precision, and such models will therefore represent a continuing source of uncertainty in simulation studies, again a matter to be assessed via the sensitivity analyses that are a central part of such studies. Some data on which to base such empirical models (mechanistic models are unlikely here) may come from pooling experience, across clinical trials, of non-consent, non-compliance, and dropout rates as a function of baseline covariates, and, for example, diverse reactions (outcomes). Some recent work defining models for compliance patterns may ultimately provide good models for simulation studies (Girard et al. 1998).

3.8 Computational Methods

The simulation plan should include descriptions of the hardware and software used for development of the models, execution of the simulation, and the programs for analysis of the simulated trial associated with each replication. Generally one should supply more details for "home grown" software than for well accepted and widely available (commercially-available and validated) software. Some special issues related to simulation are of particular note.

3.8.1 Random Number Generation

The backbone of Monte Carlo simulation is the ability to generate random numbers. It is critical that random number generation results in sufficiently "random" numbers. Random numbers can be either 'true' random numbers, which are based on actual computer hardware that usually either amplify resistor or semi-conductor diode noise, or 'pseudo-random' which are produced by a computer program. Most, if not all, statistical packages and languages incorporate pseudo-random number generators (RNGs), which use an algorithm to generate numbers that behave like 'true' random numbers sampled from a uniform

distribution. The random number generator used in any simulation project should be known to have been validated using appropriate means.

Repetition of random number sequences or other patterns may result in simulations that do not adequately represent the stochastic nature of individuals within a population (and events within the trial). Pseudo-RNGs have the disadvantage in that they are cyclical and repeat given enough calls to them. The period of a RNG is the number of calls which can be made to the RNG before the sequence repeats itself. The RNG used in a simulation should have a period at least an order of magnitude larger than the square of the number of calls to the RNG because as n, the number of function calls, approaches p true randomness decreases (Ripley 1987, L'Ecuyer 1998). Thus, RNG using a modulus near 2**31 may not have sufficient "randomness" for clinical trial simulation.

3.8.2 Simulation of Probability Densities

Once uniformly distributed random variates are generated, their values must be transformed to the appropriate probability distribution. At the very least, general simulation software packages should include the normal, log-normal, beta, and Poisson distributions with the ability to create mixture distributions from continuous variables and to truncate either discrete or continuous variables. Generation of appropriate multivariate distributions is also important.

3.8.3 Differential Equation Solvers

For description of (deterministic) time-dependent phenomena, most software makers use differential equations as their basis to have as general a program as possible, even though it is not necessary to use differential equations for all simulations. Using differential equations for IO models makes the code that implements a given (sub-)model more readable, but requires heavy computations for calculation of outputs. Linear systems may have explicit solutions, which may then be modified using linear operators to solve the problem at hand, e.g., using a one-compartment model and the superposition principle to generate a multiple dose concentration-time profile. The primary advantage of analytical equations is speed. Differential equation solvers are slower than an explicit equation solver. If the dynamic system is nonlinear (and this non-linearity bears relevance to the simulation project) Then a differential equation solver generally must be used with, as a consequence, a tremendous increase in computational requirements.

The choice of the differential equation solver depends on the problem at hand. If the integration interval is large enough, as it is with most simulations, then adequate accuracy may be obtained using Runge-Kutta or Adams methods with 4^{th} order adaptation. If the ratio of the largest to smallest rate constant is large (e.g. in pharmacokinetic compartmental modeling), or if there are slowly and rapidly varying components within the system, this stiff system requires very specific algorithms for their solution, such as

Gear's algorithm or the Livermore Solver of ODEs. Since clinical trial simulation explicitly studies the effect of variability (e.g. in PK and PD), “extreme” subjects and/or parameter values are likely to occur, so that robust integration methods will be often preferred. In any case, it is recommended to perform a preliminary check whether the ODE solver is adequate for the envisaged simulation(s).). In some cases a suitable approximation could greatly decrease the computational burden, and might be used if there is little loss from its use (i.e., some evaluation of impact is needed)

3.8.4 Computer Requirements

A fast CPU (actually “as fast as possible”) may be needed in order to perform the simulation project in a reasonable period of time, compatible with the drug development timelines. Monte Carlo simulation studies typically require large amounts of memory and storage capacity. A few trial simulations, each based on a thousand replications, can easily add up to hundred's of MB of disk storage. Adequate RAM (64 MB or greater) is typically needed to be able to manipulate data sets of this size. Application of the parsimony principle can help to avoid overwhelming available resources. The simulation plan should specify the list of those responses of the simulated study that will be stored in the simulation database.

3.9 Analyses

There are two levels of analysis. The first one operates at the level of the replication. It will describe how each individual simulated trial is to be analyzed. The second one describes how the group of simulations in the database of the simulation experiment is to be analyzed as a whole (a form of meta-analysis).
The same statistical analysis planned for the actual clinical trial should be used to analyze each replicate of an individual simulated trial. The replications of the simulated trial provide a distribution of study outcome statistics, providing insight into a probable distribution of outcomes for the actual clinical trial. At the replication level of the simulated trial, the method of statistical analysis may vary with changing study design. This should be described in the simulation plan when different study designs are investigated within a simulation project. Alternative statistical analyses allow comparison of the methods of analysis under the conditions of the simulated trials. The model used to simulate data will usually be more complex than the model proposed for the planned analysis of the actual trial. This allows evaluation of the importance of potential model misspecification in the planned analysis.
The appropriateness of statistical analytical methods for the analysis of individual simulated trials and for the meta-analysis of a group of clinical trial replications (see section 4.3.2) should be considered in the planning phase.

3.10 Critical Assessment of Simulation Results

The plan should address how the simulation results will be compared against actual trial outcomes and trial implementation. Simulation performance criteria may be very simple, such as cursory review of median & range versus anticipated values, or could be complex, such as evaluating distributional properties of simulated parameters and the types of protocol derivation that occur. Procedures to build confidence that a simulation has been properly implemented (models, distributions, sampling, etc.) should be planned as part of the simulation project.

3.11 Reporting

The plan should describe the reporting to occur following the simulations and analyses. Mockups of key tables or figures are helpful in making sure that key project objectives are well addressed.

4 EXECUTION OF THE SIMULATION PROJECT

4.1 Model Building

Clearly, a simulation project can be no better than the quality of the models it uses. Hence, considerable attention to the models (of all 3 types) is warranted, and the parsimony principle should be applied, for both the objectives of the project and the associated models. There is considerable experience with, and folklore about, model-building, but little published literature on good practices or standards. Model-building as currently practiced is an essentially inductive and hypothesis-generating activity, and has not been considered amenable to algorithmic definition. Certain practices, such as consulting subject-matter experts, are an obvious “must” at the project planning level. Such expertise should be, to some degree, already present within the trial design team.

In contrast to model-building for data analysis only, it is essential to undertake considerable model-checking before accepting a model for a simulation study, and the requirements for model performance do differ, as discussed above, from data analysis to simulation. Accordingly the next section discusses some principles of model checking or validation. .

4.2 Model Checking and Validation

Model evaluation must take into consideration the intended use of the model. At the very least, one must be able to describe anticipated future observations from the model, i.e. similar data observed under similar conditions. But models are most useful when they can be used for prediction of different data and/or under different conditions. Model evaluation should not focus on whether it is the "correct" model, but should ultimately address the predictive performance of the model. Such evaluation requires more than the usual goodness-of-fit criteria such as inspection of distributions of residual and weighted residuals and examination of standard errors of the estimates and correlations among parameter estimates. Such standard tools are insufficient to evaluate all the variance-covariance components of models involving random effects and provide little if any information about model performance when used for prediction.

Model evaluation can be divided into three parts, 1) empirical evaluation, 2) mechanistic evaluation, and 3) predictive performance. Not all parts may be relevant to a specific application. Empirical evaluation involves the question, is the model consistent with the observed data? The standard goodness-of-fit criteria partially address this question. Procedures to estimate prediction error based on the original data set may involve external validation, which involves splitting the data into learning and validation data sets and predicting the validation data from the model or, cross-validation, which is essentially repeated data splitting.

Bruno et al. (1996) demonstrated the external validation approach by prediction of parameters of interest for the validation dataset using the chosen model, followed by comparison of these predictions to a naive model (no covariates). Empirical Bayesian estimation is used to obtain the "observed" parameter estimates for the validation dataset. This approach is very useful for assessing the importance of covariates. If a model is an adequate representation of the data, it should be possible to use the model to simulate parameters and pseudo-data that are generally consistent with any prior knowledge of parameters and the observed data.

Since the use of PK/PD models for clinical trial simulation, in most cases, will require extrapolation, mechanistic evaluation may be particularly important. The model should be consistent with the underlying physiological, pharmacological and pathophysiological processes and quantities. Sensitivity analyses can be used to assess the impact of misspecified parameters and other model components and assumptions, and thereby provide some, perhaps crude, estimate of the precision of the simulation-based predictions. Considerable effort may be needed to build physiologically consistent models without making them unnecessarily complex. This effort must be done in consultation with the clinical experts who are most knowledgeable about clinical trial outcomes for a particular therapeutic intervention. This consultation may be of greatest importance when pharmacodynamic information from early trials is used to predict the actual clinical responses observed in a phase 3 efficacy/safety trial.

The ultimate test for a model is the assessment of predictive performance when the model is used to predict data from a different study or trial. This test should be employed whenever the model will be used

to extrapolate from the original study conditions and appropriate independent data are available. The type of data and the conditions under which it is collected should be as similar as possible to the planned use of the model.

Evaluation of predictive performance can be carried out at either the parameter or the observed data level. Proposed model predictions should be checked against existing data, paying particular attention to lack of fit or bias (lateral validation). Evaluation of range of validity is encouraged, as many models may be useful over a limited range but become less useful outside that range. "Spot checking" of simulated data against assumptions can help ensure correct implementation of data generation routines. Since model imperfections may lead to inaccurate or misleading simulations via propagation of errors, models to be used in simulation should be checked to assure that they are capable of generating datasets that reflect the datasets from which they are derived. The posterior predictive check (Gelman et al. 1995, 1996) for evaluating predictive performance involves Monte Carlo simulations of the original trial from which the models were derived, using the posterior distribution of population PK parameters estimated from the original trial data (or a reasonable approximation to it). The probability of any statistic derived from the original data under the fitted model can be determined from the distribution of that statistic derived from the replications of simulated trials, and provides evidence for model misfit if the probability is low. Examples of predicted characteristics might be trough concentrations at steady-state and the peak to trough concentration difference for multiple-dose pharmacokinetic data or the change in response between first and last dose for a pharmacodynamic model describing tolerance. In the context of mixed effect modeling, the posterior predictive check may have the ability to detect model misspecification in the variance-covariance model for random effects (Kowalski 1999). This is of particular importance for models used to simulate clinical trials. Inferences based on these simulations may be more sensitive to distributional assumptions. This will be especially true if the inferences are influenced by extreme observations.

For overall checking of simulation results, using graphical display is generally helpful. Visual display allows comparison of selected outcomes with prior results and a (partial) check that expectations regarding the mimicking of reality have been met

4.3 Analyses

4.3.1 Replication Analysis

The analyses planned for the actual clinical trial should ***always*** be done on the simulated data from each of the individual simulated trial replication. It may also be useful to look at alternate analyses, metrics,

variance-covariance structures, etc. to evaluate simulation strengths and weaknesses under each approach. Based on the raw data from the replications associated with a given trial design, one can generate summary values descriptive of the corresponding design, for use in analyzing the simulation project as a whole.

There may be several key statistics of interest resulting from each individual simulated trial. They might include the primary trial statistic, the primary outcome, various estimated parameters, a goodness-of-fit statistic, or any other statistics of interest , various estimated parameters, a goodness-of-fit statistic, or any other statistics of interest (e.g., such as proportion of patients responding to treatment, p-value for primary comparison, number of dropouts, or estimate of a pharmacokinetic parameter).

4.3.2 Simulation Experiment Analysis

Analysis at the level of the simulation project provides integrative and comparative insights into the group of simulations that were performed. Some measures of interest for exploration might include sensitivity, power, bias, precision, robustness, data dependence on models and design, conclusion dependence on analysis technique surrogate evaluation (e.g., agreement of surrogate with outcome), etc.

A histogram showing the distribution of a key summary statistics of interest is expected (e.g. the actual trial primary outcome variable). In addition to simple histograms, common descriptive measures of distribution (for quantitative variables) will often be useful, such as mean, median, mode, standard deviation, range, inter-quartile range, quartiles, minimum, maximum, percentiles, etc. Percent success is one appropriate measure for a pass/fail variable.

Some suggested graphic displays for consideration include: histograms (possibly smoothed), percentile summary plots, profile plots (overlaid curves), concordance plots (comparing methods, designs, etc.), scatter, contour, box-whisker, distribution diagnostic (e.g., normal plots), and possibly multi-panel of any of these types. One particularly interesting way to present premise and design/analysis joint impact is in a tabular array, with premises listed in rows and design/analysis possibilities listed in columns, with each cell providing summary information (or, even better, displaying a graphic) representing what happens at that combination.

Statistical analysis of the group of simulated trials from the data in the simulation database should be in accordance with the design of the simulation project, often using statistical methods appropriate for factorial or response surface designs. Competent statistical expertise (which will generally be present in the trial design team) is required here. The approach generally may also depend on the objectives, such as maximization of power, maximization of sensitivity, or minimization of cost. Each primary parameter estimated should include an estimate of its uncertainty. Diagnostic procedures, such as residual plots, should be also considered.

4.4 Report Contents

Guided by the principle of clarity and completeness**,** all methods and results of the simulated trials, including statistical analysis, should be interpreted and summarized as a whole in a report describing the results of the simulation project. The report should also include a statement as to whether and how the actual simulation differed from the planned simulation as stated in the simulation plan. This report should be at a level of detail sufficient to be thorough, but also to be understandable by all intended readers. This report provides a single location for decision making about the clinical trial design by incorporating all aspects in one coherent package for communication and evaluation

5 CRITICAL ASSESSMENT OF SIMULATION

Clinical trial simulation is a new and evolving tool for aiding drug development. Critical evaluation of this approach is needed to assess its value, in parallel to the increasing development and dissemination of the technology.

5.1 Prospective Evaluation

To the extent that simulation of already completed trials may be used to guide the development of this field and to expand practitioner experience, it is essential that these simulations be carried out in a completely “blinded” manner, without reference to the actual results of the completed trials. Only after evaluation of the performance of the simulation relative to the actual clinical trial outcome, should the clinical trial data be "mined" for information about why the simulations may or may not have been a reasonable representation of the actual trial being considered.

5.2 Retrospective Evaluation

The actual prediction of future trials based on simulations is, of course, self-blinding because those responsible for the simulation cannot know the outcome of the future trial. It is important, though, to capture information about simulation performance and the reasons for general "success" or "failure", once the actual trial is completed. At best, clinical trial simulation can provide an intelligent estimate of the range or distribution of likely outcomes based on available data. The outcome of a given real trial

represents only one realization of the trial and as such may or may not fall in the range of typical or usual outcomes. Also, it may happen that a trial simulation based on an imperfect model may still have provided the right answer regarding the choice of clinical trial design.

5.3 Cumulative Evaluation

Any hope of assessing the overall value of clinical trial simulation will come from cumulative experience. Accumulation of data on protocol execution deviations and on other aspects of clinical trials (e.g., across center differences, geographical differences in placebo effects, etc.) and their integration into models for inclusion in clinical trial simulation are among many future challenges to be met in order to construct simulated trials that better represent the actual trial experience. Therefore, it is vitally important to maintain an ongoing compilation of experiences and "lessons learned" in clinical trial simulation from all sources.

6 REFERENCES

Box GEP, Hunter SJ, Hunter WG. Statistics for Experimenters : An Introduction to Design, Data Analysis, and Model Building, 1978, Wiley

Bruno R, Vivler N, Vergniol JC, De Phillips SL, Montay G, Sheiner LB. A Population Pharmacokinetic Model for Docetaxel (Taxotere): Model Building and Validation. Journal of Pharmacokinetics and Biopharmaceutics 1996; 24:153-172

CDDS Conference Proceedings, Modeling and Simulation of Clinical Trials in Drug Development and Regulation, Herdon, VA, November 10-11, 1997

ECPM/CDDS Workshop Proceedings, Frontiers in Drug Development: Computer Simulation and Modelling, Basel, Switzerland, October 18, 1996

FDA Guidance for Industry, "Population Pharmacokinetics". Feb 1999. U.S. DHHS, FDA, Center for Drug Evaluation and Research, Rockville, MD, USA. http://www.fda.gov/cder/guidance/1852fnl.pdf

Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis. Chapman & Hall, London, 1995

Gelman A, Meng XL, Stern H. Posterior Predictive Assessment of Model Fitness via Realized Discrepancies. Statistica Sinica 1996; 6: 733-807

Girard P, Blaschke TF, Kastrissios H, Sheiner LB. A Markov Mixed Effect Regression Model for Drug Compliance. Statistics in Medicine 1998; 17: 2313-2333

Hale M, Gillespie WR, Gupta S, Tuk B, Holford NHG. Clinical Trial Simulation as a Tool for Increased Drug Development Efficiency. Applied Clinical Trials 1996; 5:35-40

Johnson SCD. The role of simulation in the management of research: what can the pharmaceutical industry learn from the aerospace industry? Drug Information Journal 1998; 32:961-969

Kowalski K, Table 12, CDDS M&S Workshop Proceedings, February 1999

Krall RL, Engleman KH, Ko HC, Peck CC. Clinical Trial Modeling and Simulation - Work in Progress. Drug Information Journal 1998; 32:971-976

L'Ecuyer P. Random Number Generation in Handbook on Simulation (Ed: Banks J). John Wiley and Sons, New York, 1998

Peck C, Desjardins, R. Simulation of Clinical Trials - Encouragement and Cautions. Applied Clinical Trials 1996; 5:30-32

Peck CC: Drug development: Improving the process. Food and Drug Law Journal 1997; 52(2):163-167

Peck C. Defining Modeling and Simulation. pp 15-18 in CDDS Conference Proceedings, Modeling and Simulation of Clinical Trials in Drug Development and Regulation, Herndon, VA, November 10-11, 1997

Rubin DB. Multiple imputation after 18+ Years. Journal of the American Statistical Association 1996; 91: 473-489

Ripley BD. Stochastic Simulation. John Wiley and Sons, New York, 1987

Sacks J, Schiller SB, Welch WJ. Designs for computer experiments. Technometrics 1989; 31: 41-47

Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. Statistical Science 1989; 4: 409-435

The Third National Health and Nutrition Examination Survey 1988-94 (NHANES III) US Centers for Disease Control and Prevention Division of Health Examination Statistics. National Center for Health Statistics (NCHS). http://www.cehn.org/cehn/resourceguide/nhanes.html

Urquhart, J. Measuring compliance -- how can it be done? Plenary Lecture, International Workshop on Validated Outcome Measurements in Pharmaceutical Care. Danish College of Pharmacy and the Pharmaceutical Care Network Europe, Hillerod, Denmark, January 28, 1999.

Welch WJ, Buck RJ, Sacks J, Wynn HP, Mitchell TJ, Morris MD. Screening, predicting, and computer experiments. Technometrics 1992; 34: 15-25