

## Using SAS® for Modeling and Simulation in Drug Development – A Review and Assessment of Some Available Tools

Melvin Munsaka, TGRD, Deerfield, Illinois  
Michael Carniello, TGRD, Deerfield, Illinois

### ABSTRACT

It is widely recognized that modeling and simulation is a valuable tool that can be used to help understand various considerations in clinical trials, such as the evaluation of merits of different designs and potential efficacy and safety outcome scenarios. Together with available scientific data, this can be used to help assess and evaluate potential risks and outcomes of trials before making huge investments. Simulation and modeling can also help provide guidance in decision making by quantifying and evaluating potential choices such as the go-no-go decision, in the face of a multitude of uncertainties. Modeling and simulation can in addition be used to help define, test, compare, and optimize study designs, including specific complex design scenarios that are useful for facilitating a better understanding of study outcomes. In this survey, we review and assess the use of SAS® as a tool in clinical trial modeling and simulation in drug development.

### INTRODUCTION

Modeling and simulation within the context of clinical trials has become an important tool used in addressing a variety of considerations and is now a widely used component for improving the efficiency of drug development. It is recognized as a powerful tool that can be used to help drug development more informed and more efficient. It has received prominence as one of the tools that are being advocated by both the pharmaceutical industry and regulatory agencies as a new approach to drug development.

Modeling and simulation can be applied throughout the many phases of drug development. Segreti (2001) *et al* noted that with readily available desktop computational power, the statistical methods used in analyzing clinical trials have become more sophisticated with simulation being a routine tool used to address study design questions, including those raised by FDA reviewers. Also, as observed by Netorov *et al* (2001), modeling and simulation are well recognized powerful tools that enable effective implementation of the so called learn-and-confirm paradigm in drug development. Chang (2010) discusses a variety of considerations of Monte Carlo simulation in the pharmaceutical industry.

Modeling and simulations can be used to study and compare potential inferential methods, such as exploring model properties, estimates, and sensitivity to uncontrollable inputs such as dropouts and missing data. Modeling and simulation can be used to address drug development questions that typically come up such as dose optimization, potential for stopping development early due to safety concerns or overwhelming efficacy evidence, anticipated efficacy and safety outcomes, identification of subgroups with a higher affinity to certain adverse events or those more likely to elicit a favorable response thereby allowing for more informed risk-benefit assessment.

The planned analysis of a trial under a specified study design assumes that it will be executed according to the protocol. However, complete adherence to the protocol by all subjects is rare; thus safety and efficacy assessment based on what was initially assumed no longer holds. Additionally, deviations from the protocol can lead to the failure of a study to address the stated objectives. Modeling and simulation offer one way of investigating how potential deviations can affect the trial and also identify potential weaknesses or limitations of a proposed study design. This can in turn be used to develop workable plans on how to avoid, minimize, and/or resolve these limitations to increase study robustness prior to execution.

Simulations can be used for defining and testing interactive drug models, exploring and communicating study design attributes, and performing analyses about precision and accuracy of future endpoint estimates. A good simulation incorporates scientific knowledge to help the entire project team communicate and test ideas and to plan effective trials for every phase of clinical development.

Another important use of modeling and simulation is that of using scenario trials or mock up trials regarding a plausible profile of the safety and efficacy of a treatment taking into account both aspects that are favorable and non favorable to the study. This would require input from various functional areas to explore a series of scenarios and assumptions and different trial designs, and to explore alternative ways of analyzing the simulated data. This can in turn help inform the team about the impact of alternative designs and what analysis methods are appropriate for the anticipated outcomes of the study. Simulations can also be used to explore situations that would be ethically, financially, or logistically impossible to carry out in practice.

In recognition of the immense potential that modeling and simulation has for making drug development more efficient, many pharmaceutical companies have setup departments and assembled various software tools specifically to focus on modeling and simulation (see, for example, Buchheit, 2007). These tools are increasingly used to evaluate the merits of study designs in the context of a drug development program as a whole. A review of the literature suggests that a wide variety of software

is used for modeling and simulation. In this survey, we focused on identifying, reviewing, and assessing how SAS is being used as a tool in clinical trial modeling and simulation within drug development.

## **DEFINING MODELING AND SIMULATION**

Interpretations of modeling and simulations can differ among informed researchers. In this survey, a model is defined as a mathematical relationship among pieces of data. A model is a conceptual tool for translating often complex, real-world subject matter into a simplified form. Modeling should account for both deterministic and stochastic considerations. Within the clinical trial setting, it should allow for synthesis of data within a study and/or across inter-related studies and facilitate an examination of what can be expected in planned clinical trials, thereby optimizing clinical study design and planning.

Simulation is the process of using a mathematical model that mimics a real-world situation to conduct experiments in order to describe, explain, investigate, and predict the behavior of that situation. It includes the process of simulating data based on an assumed model. A simulation allows for visualization and communication of system behavior and exploration of model properties, model structure, model parameters, and model validity. Burton *et al* (2006) have provided a discussion of the key statistical considerations when performing simulations in clinical trial settings. The considerations include random number generation, methods for generating data, scenarios for investigation, and methods for evaluation, estimates, number of simulations required, and evaluation of statistical methods for different scenarios.

## **SPECIFIC USES OF MODELING AND SIMULATION EXAMINED**

Our methodology consisted of searching the internet for key phrases such as “SAS modeling simulation”. We reviewed typical SAS related publications such as Proceedings of PharmaSUG, NESUG, SESUG, and SAS Global Forum. We extended the search to likely journals, such as the Pharmaceutical Users Software Exchange (PhUSE) and other journals that are likely to discuss modeling and simulation within the clinical trial setting, including statistical methodology journals. For discussion purposes, we classified our findings into the following areas: enrollment, sample size estimation and re-estimation, clinical data simulation, assessment of statistical methods and missing data, complex modeling and design, bootstrapping and Monte Carlo Markov Chain, and pharmacokinetics and pharmacodynamics modeling.

Some of these areas may be inter-related and hence by no means mutually exclusive. Further, the search we performed is by no means exhaustive nor do we make any claim on the completeness of the search regarding the literature on clinical trial modeling and simulation using SAS. However, we do believe that what we present is useful as a starting point for anyone who is interested in looking into various ways that SAS is being used for modeling and simulation in clinical trials.

## **ENROLLMENT**

To execute a clinical trial well requires skilled staff in a variety of positions doing a variety of complicated tasks. Even if all that work is done well, there is a key factor that is outside the control of the staff involved, namely, enrollment rates or recruitment of subjects. The allocation of sufficient time for participant recruitment is one of the fundamental aspects in planning a clinical trial. Enrollment is logistically hard, if not impossible, to control and predict with certainty. The best that sponsors can do, typically, is predict enrollment patterns and rates, and use that information for planning, budgeting and the like. Enrollment, however, typically cannot be presented analytically – in some closed-form fashion – and so simulation is needed to help with providing information on how enrollment may proceed. Abbas, Rovira, and Casanovas (2007) describe different simulation models under continuous and discrete time assumptions for the design of recruitment in clinical trials. They propose Monte Carlo simulation Markov models which could help optimize the recruitment process and via a continuous time simulation that could minimize the duration of the recruitment and the total duration of the trial.

Monti (2004) considers trials which have many sites. In these trials, it's typically desired that the total enrollment target (say, “N”) is divided evenly among the sites (if K sites, then the target would be each site having N/K subjects). However, that is rarely achieved in real life, so some sort imbalance is expected. The article used simulation to compare the magnitude of the treatment imbalance if randomization were performed in permuted blocks of the target sample size. The SAS procedure PROC PLAN was used to generate assignments in the simulation.

Carter (2004) considered the issue of determining an optimal period of time for participant recruitment. Assuming a Poisson process and via a simulation, he showed how one can empirically estimate the distribution of accrual time in clinical trials. The simulation study used provides the means to generate an empirical probability density function for the recruitment time based on time-dependent changes in the accrual rate. The simulation study presented was performed using the SAS. Accrual rates were also examined by Carter *et al* (2005). In that report, three approaches to estimating the accrual period of a clinical trial were examined (unconditional, conditional and Poisson Process).

The SAS function RANEXP plays a key role for simulating enrollment rates as described by in an article by DePuy (2009). Instead of a uniform distribution approach to subjects entering sites, an exponential distribution approach was considered (thus the use of this SAS function). Predictions and confidence intervals were generated as part of simulation results.

In the articles by Chen *et al* (2010) and Dowlamn *et al* (2004), the practical sides of subject enrollment are discussed with respect to clinical supplies – part of a whole “supply chain” approach to enrollment. The authors in each case do use Monte Carlo approaches to simulate clinical supplies throughout the study course. Being able to accurately model supplies can mean cost savings and assure the trial will continue as planned. Although SAS is not explicitly referenced as part of the simulation work, those approaches are easily transferrable to the SAS environment.

In addition to being critical to estimating how long a study will last, enrollment is a crucial factor if a study has a Data Safety Monitoring Board (DSMB). Ying and Heitjan (2008) and Ying (2010) simulates subject enrollment as following a Weibull distribution (the Weibull being sort of the exponential distribution with an additional “shape” parameter). Again, SAS is not explicitly used, but it’s easy to see how the ideas in the article can be adapted to the SAS language.

## **SAMPLE SIZE ESTIMATION AND RE-ESTIMATION**

When setting up a clinical trial, it’s key that the “proper” number of subjects are enrolled. “Proper” means many things in this context: enough subjects to prove safety; enough subjects to prove efficacy; enough subjects to meet requirements of planned statistical analysis models. As such, sample size estimation is key in clinical trials, then, and not something which can be easily derived in a straightforward analytical fashion. Simulation is frequently used to examine the tradeoffs of sample size against the perceived needs of the trial.

Within the context of Phase 1 studies, sample size estimation has been discussed by Zhou *et al* (2010, 2006, 2004). Zhou *et al* (2010) offer a SAS macro to determine power when the method of “Schuirmann’s Two One-Sided Tests” approach is used. Zhou *et al* (2006) also offer a SAS macro to determine power via simulation when considering dose proportionality. In some Phase 1 settings, higher order crossover trials are often performed as they allow for appropriate estimation of carryover effect. A SAS macro is presented by Zhou *et al* (2005) for computing empirical power in such trials.

The use of Proc IML procedure is discussed by Peszek (2004) as a basis for sample size calculations using statistical simulation. This approach is different than the typical simulation approach in SAS work which mostly uses a macro routine approach couple with DATA steps.

Wanga *et al* (2008) discuss a two-stage adaptive design; they consider an interim analysis as an “internal pilot study” and use results to determine sample size. They offer a SAS macro to simulate the process of sample size adjustment and a randomization test. Zellner, Zellner, and Keller (2001) also discuss the idea of an “internal pilot” study; using the results of the first subjects to get some idea of variability so that the “real” sample size can be derived better. A SAS macro which implements the EM algorithm is described.

Gastanagaa (2006) discuss a SAS macro which offers power calculations for generalized linear model in observation longitudinal studies.

In Phase 2 and 3 oncology studies, the endpoint is typically PFS – progression-free survival (meaning disease progression or death – whichever comes first). The logrank test is typically used for statistical analysis. Chang (2006) discusses a SAS macro to determine sample size adjustment for interval-censoring and taking into account the hazard ratio of survival.

The therapeutic area of chronic pulmonary diseases has complicated endpoints regarding lung density measurements (and longitudinal covariates). The associated complex statistical analysis means sample size is difficult to estimate; Chen *et al* (2008) discusses a SAS macro which uses an empirical approach.

Tukey’s trend test is frequently to help determining the “No Statistical Significance of Trend” dose, in a dose-response study. Duncan (2000) presented a SAS program which uses simulation to estimate sample size and power tradeoffs when using this test.

Correlated binary outcomes are considered by Dang *et al* (2008); they use a complex statistical analysis using PROC GLIMMIX which can be used for the analysis but closed-form or analytical solutions to sample size estimate won’t work. A SAS macro is described which computes power and sample sizes.

## **CLINICAL DATA SIMULATION**

In a typical clinical trial, there are hundreds of responses or observation made, both safety and efficacy, measured for each subject over the course of the study. These are collected, cleaned and subsequently analyzed to help establish safety and efficacy claims. There are clear cost benefits to be had if sponsoring companies could simulate all the responses of subjects in a study without having to conduct the actual study! There is also real patient benefit; such an approach would put no subjects at any physical risk.

A single set of simulated responses has been addressed by Cary (1997) using a straightforward approach using the uniform random distribution in a DATA step to generating BMI data. Generating data from a variety of distributions is demonstrated in a SAS macro described by Zhang (2007) using the SAS function “RAND” as the basis.

The focus of the article by Cheng (2004) is the application of survival analysis output in SAS using ODS methods; however, a by-product of this is a technique to generating “time-to-event” data using binary and exponential random number generation.

As mentioned earlier, the natural inclination of simulation approaches of SAS users is reflected in the majority of the papers we surveyed: macro routines holding DATA steps. Cassell (2007) offers a fresh method using BY-group processing to do simulations and re-sampling plans.

Welch and Slater (2010) take a very comprehensive approach to simulating all sorts of clinical data; this work was done to have available datasets ready to pilot various display programs. The article considers a hierarchy of data types for generation: first considering character or numeric, then continuous or categorical – and then dichotomous or polytomous outcomes. A SAS macro routine was used to produce the simulated data.

The above efforts represent a big step toward the concept of generating the whole range of data collection for a subject in a (simulated) trial. We also found several good references for generating (pseudo) random data - but not specifically geared to clinical trial work. For example, Han (2006) discussed the issue of generating multivariate normal data using PROC IML. Downer (2000) discusses using PROC SIM2D to simulate normal spatial data with a variety of covariance structures. The example provided reflects the agriculture industry but could be applied to the field of clinical trial simulation. Finally Erdman uses the PROC MODEL procedure in SAS/ETS to do business planning and forecasting – in a pure financial sense. Unlike other approaches discussed, Monte Carlo simulation is embedded as an option within PROC MODEL statements to use if needed.

Occasionally, it is necessary to consider generic clinical trial data that allows for realistic clinical trials data sets. Such a model should allow for general correlation structures for endpoint-time point data and non-normal distributions, including time-to-event settings while allowing for certain clinical common experiences, such as patient dropout and noncompliance. A generic template of this nature typically demands computationally efficient algorithms for clinical trials simulations. Westfall *et al* (2008, 2010) have discussed such a generic template using a grid-enabled SAS-based system.

Pharmacogenetics deals with identification of genetic variations influencing drug response. These variations are identified as markers for efficacy and safety of the drug and hence can predict both these aspects. A reasonable understanding of the complexity of pharmacogenetic influence on safety and efficacy can help in predicting the requirements for a successful study. For example, knowledge of pharmacogenetics can help in reducing adverse drug reactions, developing adaptive designs, and reducing attrition rate of drugs during development. Tangellamudi (2007) considered clinical trial simulation of pharmacogenetic data. SAS was used to develop a tool to simulate a selected gene-gene interaction model. Though not aimed at accurately modeling a gene-gene treatment interaction and addressing all the different kinds of possible interactions, the tool demonstrates the difficulty and practicality of undertaking pharmacogenetic studies under certain assumptions. The tool can be modified and developed to address a majority of interaction patterns and genetic variables to more accurately predict the study parameters.

## **ASSESSMENT OF STATISTICAL METHODS AND MISSING DATA**

There are many potential uses of modeling and simulation directed at the objective of comparing statistical methods of analysis or aimed at how best to deal with incomplete or missing data, which is a common phenomena in clinical trials. SAS is a widely used tool in these settings. SAS can be used to provide insight as to the appropriate model as well as providing some idea of how missing data can affect the results, and how to handle incomplete or missing data. Below we cite a few examples where SAS has been used for this purpose.

Khutoryansky (2001, 2002) considered the use of SAS for imputing missing longitudinal data. A simulation was used to generate longitudinal data which were used to compare some methods of imputing missing data, including last observation carried forward (LOCF), mixed effects model repeated measures method using PROC MIXED, incremental methods of imputing, and imputation methods available in PROC MI. Khutoryansky (2003) also considered imputing data in the case of longitudinal data of two correlated endpoints. Simulated longitudinal data, assumed to follow a bivariate Markov process, was used to compare some methods of imputing missing data in this setting using PROC MIXED and PROC MI. The dropout mechanism was based on individual previously observed values of one of the endpoints. Khutoryansky and Huang (2004) also discussed imputation of missing data with informative dropout. They used a simulation to generate data resembling HbA1c, an efficacy endpoint commonly used in diabetes studies, to compare methods for imputing data including those available in PROC MI.

Mueller-Cohrs (2006) looked at different methods for analyzing normally distributed data from a simple two-period, two-treatment crossover trial in those situations where some data from one of the two periods are missing; this is a common phenomena in cross-over trials. A simulation was used to investigate the size and the power of the statistical tests under various settings. The simulation took into account the number and pattern of the available data and the covariance structure of the variables to resemble potential real-world settings.

Litaker and Ferris (2004) considered the situation where there are two potential methods for analysis of correlated proportions, an analysis of variance (ANOVA) model and generalized estimating equations (GEE). They used simulated data to compare the performance of PROC GLM and PROC GENMOD. The simulation was based on a clinical study that used three diagnostic techniques, validated with biopsy. The simulated observations represent agreement or disagreement with biopsy results. Simulations included the null and two alternative scenarios, each with complete data and with 5% and 10% missing data. The binomial distribution and logit link function were used in the GEE analysis. They concluded that the methods were comparable.

Another example of where SAS has been used to examine statistical properties of statistical methods has been considered by Patterson and Jones (2004) who looked at statistical properties of the method of moments and restricted maximum likelihood in replicate designs and the application of these techniques in the assessment of average, individual and population bioequivalence.

## COMPLEX MODELING AND DESIGN

A clinical trial is a process with measurements or observations that are based on complex disease processes. Such complex processes can occur when one takes into account real world situations where, for example, an outcome can be related to other outcomes. Another setting is one where it may be of interest to examine disease progression, either by modeling a trend or using multistate models which mimic how a patient may progress through possible states of a disease, some of which may be recurrent. SAS has been used in these settings to help elucidate a better understanding of the process at hand.

Paes, Carlos, and de Lima (2004) used SAS to study event history analyses where the event of interest is non-fatal and could occur more than once for each subject. In this kind of setting, the experimental units can move between states over time, and it is possible to estimate the corresponding transition probabilities using regression models which incorporate covariates. For this purpose, they used a SAS macro to estimate the transition probabilities using three approaches for the regression modeling. The macro developed is flexible and allows inclusion of covariates with output that includes the predicted transition probabilities as a function of time.

Zhang and Zhang (2010) developed SAS macros to estimate cumulative incidence curves under the proportional hazards model allowing for computation of treatment-specific cumulative incidence functions while adjusting for potentially imbalanced prognostic factors among treatment groups. The macros they developed allow computation of standard errors for the direct adjusted cumulative incidence estimates and standard errors for the differences of adjusted cumulative incidence functions between any two treatments. They applied the macros to bone marrow transplant data.

Rosthøj, Andersen, and Abildstrom (2004) discuss SAS macros for estimating the cumulative incidences functions based on the Cox regression model for competing risks survival data. Two SAS macros were developed, one for estimating the cumulative incidences and the other for estimating cumulative incidences and the variances of the estimated cumulative incidences.

Hui-Min, Ming-Fang, and Chen (2004) developed a SAS macro for non-homogeneous Markov process modeling of multi-state disease progression. In general, modeling multi-state disease processes for cancer or chronic disease is a challenging task. The macro they developed is flexible in the sense that it can estimate the transition parameters for both homogeneous and non-homogeneous models. Covariates can also be incorporated using the proportional hazards form. They applied the macro to a three-state disease model for the progression of colorectal cancer from normal (disease free), to adenoma (pre-invasive disease), and finally to invasive carcinoma, with or without adjusting for covariates. The macro can be generalized to  $k$ -state models with  $s$ -covariates.

Abnormal laboratory data from clinical trials are considered precursors or confirmatory factors of potential organ dysfunction. Typically, the analyses of these data focus on the incidence of extreme values relative to normal ranges and/or change relative to baseline. Another approach is to focus on the worst values from each subject. However, these approaches ignore the dynamics of the measurements over time and rarely is there an attempt to address this from a modeling perspective. Rosenkranz (2009) considered modeling laboratory data using PROC NL MIXED based on the Ornstein-Uhlenbeck process which has some properties that are suitable for this purpose. These models can account for potential systematic effects relevant for all or a group of subjects as well as the random variation between subjects with the objective of quantifying treatment effects.

The use of specific methods for planning of a clinical trial is critical for the efficient execution and methods have been discussed in the literature. One such method is that of group sequential trials; the benefits are generally acknowledged with acceptance from the regulators when used properly. Whitehead (2010) discusses an implementation of a wide range of group sequential designs using two accessible SAS functions. These include PROBNORM, a standard function and the other, SEQ, which is part of PROC IML. The discussion focuses on the essentials of group sequential designs and reveals how straightforward it can be implemented within SAS.

Another method is the continual re-assessment (CRM) method introduced by O'Quigley *et al* (1996) for dose escalation to address concerns noted on the conduct of certain trials, including ineffectiveness of treatment at low dose, toxicity, lack of knowledge of the dose-response relationship, and efficient study design in cases where the sample size is small. This method and its modifications have been discussed in the literature. Adam (2010) considered an implementation of the CRM using SAS in a Phase I oncology trial in patients with advanced solid tumors with the primary objective of determining the MTD. The analysis used the two-parameter logistic regression model. Ishizuka (2001) also discussed an implementation of the CRM in SAS using the QUAD call within PROC IML. He also implemented some graphical output using PROC GPLOT. A specific software called Bayesian ADEPT that implements the CRM and runs within a SAS using BASE, STAT, and GRAPH is also available commercially (see Bedding, 2003).

An implementation of CRM method for time-to-event (TITE-CRM) as described by Cheung and Chappell (2000) was also implemented in SAS (UMICH, 2008). The program developed makes all of the calculations necessary to determine dose allocation of a trial in real-time and also includes the ability to simulate complete trials in order to evaluate the operating characteristics of a design prior to implementation.

## **BOOTSTRAPPING AND MONTE CARLO MARKOV CHAIN**

In clinical trials, it is common that both observed and derived measurements follow unknown distributions. It is also necessary to give some indication of the accuracy of efficacy and safety estimates. Without knowledge of the distribution, parametric approaches cannot be relied on; this calls for the use alternative estimation methods. Many analysis methods assume normality; when this assumption is violated, inference cannot be reliable. In these cases the use of computer intensive estimation methods such as bootstrapping, a re-sampling technique, is used to obtain refined estimates and particularly precision measures. The most common application of the bootstrapping technique may be in obtaining the confidence interval for an estimator where the typical way of obtaining the confidence interval through the standard error approach is impossible or difficult. Barker (2005), Miller (2004) and Cole (1999) give useful introductions of bootstrapping within the SAS context. Bootstrapping is also implemented in some of the SAS procedures, such as PROC MI and PROC MULTTEST.

Bayesian methods have become increasingly popular in clinical trial. They allow for the incorporation of prior information, and for simultaneous inference using data from multiple studies. Bayesian inference is also useful for complex hierarchical models. The SAS MCMC (see Chen, 2009) procedure allows for a variety of Bayesian statistical models that can be used in the clinical trial setting. The procedure is based on Markov Chain Monte Carlo (MCMC) methods and performs posterior sampling and statistical inference for Bayesian parametric models. It can fit single-level or multilevel models which can take various forms, from linear to nonlinear models, by using standard or nonstandard distributions.

Wang, Carpenter and Kepler (2006) considered using SAS to conduct nonparametric residual bootstrap multilevel modeling with a small number of groups. In this particular setting, one has to deal with the issue of sparse data at high levels in multilevel modeling. In particular, it is common to encounter data with a relatively small number of observations at the higher levels. As a result and the potential for non-normality of the residuals, parameter estimates, particularly the variance components and standard errors of parameter estimates at the group level, may be biased, resulting in unreliable inference. The bootstrap method allows for the estimation the standard errors of the parameter estimates for significance testing.

Tanriverdio and Ring (2007) considered implementation of bootstrapping in a replicate bioequivalence study with a highly variable drug where there are challenges in the establishing bioequivalence using the usual rules. For highly variable drugs, it is difficult to apply the usual bioequivalence methods as a large sample size is needed to show that the confidence interval of the ratio of the average lies completely within the usual bioequivalence bounds. Tanriverdio and Ring used the scaled average bioequivalence-method which is technically more demanding and the resulting statistics cannot be assumed to be normally distributed. They illustrated how SAS can be used to address the normality question using the bootstrap method.

Gemperli (2010) considered the use of MCMC methods in the analysis of safety data using Bayesian simulation methods. The method that he applied made use of SAS to apply the hierarchical model proposed by Berry and Berry (2004) for the analysis of frequency counts of adverse events with one active treatment and one control group. For parameter estimation, a simulation-based Bayesian approach was used. They made use of PROC MCMC which offers the required features to implement the Berry-Berry model within SAS.

Patterson and Yeh (2007) also discussed the use of SAS for MCMC simulation in the context of clinical trials to perform Bayesian statistical analysis. They simulated hypothetical clinical trial efficacy data and used MCMC methods for missing data imputation, model parameter simulation, and model diagnostics.

## **PHARMACOKINETIC AND PHARMACODYNAMIC MODELING**

Pharmacokinetics (PK) is study of the kinetics of absorption, distribution, metabolism, and excretion of drugs. The focus of pharmacodynamics (PD) on the other hand is on the pharmacologic response or effect at the receptor site. Both pharmacokinetics and pharmacodynamics play an important role in drug development and are critical in determining the optimum dosage regimen and form taking into account a balance between efficacy and safety. It is thus important that they are both addressed adequately in a drug development program. Another important consideration is that of population pharmacokinetics and pharmacodynamics which can be done both at early and late stages of clinical development. The population-based approach is used to identify and interpret the demographic, genetic, physiologic and pathologic factors on pharmacokinetics and pharmacodynamics. Below, we highlight how SAS has been used for PK and PK modeling and simulation.

Russek-Cohen, Martinez, and Nevius (2005) developed programs using PROC IML for simulating pharmacokinetic datasets that reflect the various kinetic, population, and study design characteristics that complicate the bioequivalence evaluation. The simulation made use of the one and two compartment models. Erdman and Morelock (1996) considered a study of kinetics focusing on the estimation and simulation of systems of first-order differential equations using PROC IML.

Su *et al* (2009) considered the use of SAS to fit compartmental models under multiple dosing using PROC NL MIXED. The program they wrote considered construction of mean functions for fitting pharmacokinetic data from multiple dose studies. One-compartment models for oral dose administration were considered. Shen and Lu (2007) also used SAS for population pharmacokinetics using nonlinear mixed effects modeling. They discussed how this can be used to estimate the typical PK and/or PD parameters and to identify sources of variability (covariates) that explain differences in the parameters between patients, and quantify the impact of these covariates. They used the NL MIXED procedure.

Hoffman *et al* (2005) considered using nonlinear mixed effects modeling for estimation of steady state attainment. They used SAS for the assessment of the time needed to attain steady state based on a nonlinear model and compared this to ANOVA-based methods. The method used was a nonlinear mixed effects modeling approach for estimation of steady state attainment, based on fitting a simple nonlinear mixed model to observed trough plasma concentrations. Method performance of the nonlinear mixed modeling versus the ANOVA-based approaches was done via simulation.

Wang (2005) considered using SAS in modeling the relationship between concentration and response which is challenging when the drug concentration and response measurements are not taken at the same time and/or hysteresis occurs between the response and the concentration. To help address this concern, an algorithm for a semi-parametric approach to fitting nonlinear mixed PK/PD was proposed which included an effect compartment model using linear interpolation and extrapolation for concentration data. This approach is independent of the PK model and the algorithm was implemented using SAS PROC NL MIXED. The properties of the approach were evaluated using a simulation and was also applied to a study of the PK/PD relationship between insulin and glucose levels.

Unless all of a drug is eliminated during each dosing interval, the plasma concentrations within a dosing interval will increase until steady state is reached as result of drug accumulation. Panebianco and Maes (2011) considered the estimation of time to steady state using the effective rate of drug accumulation. They assumed a one-compartmental model with first-order absorption and first-order elimination based on a deterministic formula for plasma concentration at time  $t$  within a dosing interval. They estimated time to steady state based on multiple dose area under the plasma concentration–time curve. Several point estimates and confidence intervals for time to 90% of steady state were compared. SAS was used for the data simulations and analyses.

## CONCLUSION

We have attempted to give a snapshot of the use of SAS in modeling and simulation within clinical trials. The discussion has shown that SAS is widely used in various considerations in clinical trial modeling and simulation. We hope this paper provides a good start for interested SAS users to see how their peers in industry and academia have used SAS for modeling and simulation work. Performing modeling and simulation within SAS provides an opportunity to utilize the statistical capabilities of this software platform. Further, SAS has the same capabilities as other software for performing modeling and simulation from the simplest to the most complex settings (see, for example, Smith, 2003). In his assessment, O'Kelly (2009) concluded that SAS is somewhat faster than R for simulations where no generalized linear modeling is involved and that SAS is considerably faster than R where generalized linear modeling is involved in simulations. There is readily available informative literature some of it included in this paper and user developed SAS programs and macros in the public domain for modeling and simulation.

There remains opportunity, we feel, in the area of clinical trial modeling and simulation. Instead of pieces of a trial, it's feasible that a whole trial can be conducted "virtually" – saving sponsor time and money, and eliminating clinical testing risk and making clinical trials more efficient.

## REFERENCES

- Abbas, I., Rovira, J., and Casanovas, J. (2007). Clinical trial optimization: Monte Carlo simulation Markov model for planning clinical trials recruitment. *Contemporary Clinical Trials*, **28**, 220-231.
- Barker, N. (2005). A Practical introduction to the Bootstrap using the SAS System. PhUSE, Paper PK02.
- Barrett, J.S., Jayaraman, B., Patel, D., Jeffrey, M., and Skolnik, J.M. (2008). A SAS-based solution to evaluate study design efficiency of phase I pediatric oncology trials via discrete event simulation. *Comp. Meth. Prog. Biom*, **90**, 240–250.
- Bedding, A. (2003). Bayesian ADEPT assisted decision making in early phase trials. *Pharmaceut. Statist.*, **2**, 219–221.
- Berry, SM and Berry, DA (2004). Accounting for multiplicities in assessing drug safety: a three-level hierarchical mixture model. *Biometrics*, **60**, 418-426.
- Buchheit, V. (2007). A dedicated programming group working in a pharmaceutical modeling and organization. PhUSE, Paper PD03.
- Burton, A., Altman, D. G., Royston, P., and Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statist. Medic.*, **25**, 4279-4292.
- Cassell, D. L. (2010). BootstrapMania!: Re-Sampling the SAS® way. SAS Global Forum, Paper 268-2010.
- Cassell, D. L. (2007). Don't Be Loopy: Re-Sampling and Simulation the SAS® Way. SAS Global Forum, Paper 183-2007.
- Carter, R. E. (2004). Application of stochastic processes to participant recruitment in clinical trials. *Controlled Clinical Trials*, **25**, 429-436.
- Carter, R. E., Sonne, S. C., and Brady, K. T. (2005). Practical considerations for estimating clinical trial accrual periods: application to a multi-center effectiveness study. *BMC Med. Res. Meth.*, **5**:11.

- Cary, A.J.L. (1997). Generating data with the SAS Dataset. SUGI 22, Paper 74.
- Chang, M. (2010). *Monte Carlo Simulations for the Pharmaceutical Industry*. CRC Press.
- Chang, C-K. (2006). Sample size calculation and timeline estimate for progression-free survival. PharmaSUG 2006, Paper SP07.
- Chen, F. (2009). Bayesian modeling using the MCMC Procedure. SAS Global Forum, Paper 257-2009.
- Chen, J., Stock, S., Deng, C. (2008). Sample Size estimation through simulation of a random coefficient model by using SAS. PharmaSUG, Paper SP03-2008.
- Chen, Y., Mockus, L., Orcun, S., and Reklaitis, G. (2010). Simulation-based optimization approach to clinical trial supply chain management. 20th European Symposium on Computer Aided Process Engineering - ESCAPE20.
- Cheng, W. (2004). Helpful Undocumented Features in SAS®. SUGI 29, paper 040-29.
- Cheung, Y.K. and Chappell, R. (2000). Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics*, **56**, 1177-1182.
- Cole, S. R. (1999). Simple bootstrap statistical inference using the SAS system. *Comp. Meth. Prog. Biomed.*, **60**, 79-82.
- Corbiere, F. and Joly, P. (2007). A SAS macro for parametric and semiparametric mixture cure models. *Comp. Meth. Prog. Biom.*, **85**, 2007, 173–180.
- Dang, Q., Mazumdar, S., and Houck, P. R. (2008). Sample size and power calculations based on generalized linear mixed models with correlated binary outcomes. *Comp. Meth. Prog. Biom.*, **91**, 122–127.
- Davis, A. (2009). Adhoc Analysis of Site By Treatment Interaction. PharmaSUG, Paper SP07-2009.
- DePuy, V. (2009). Enrollment simulation in clinical trials. SESUG 2009, Paper SD-0009.
- Dowlamn, N., Lang, M., McEntegart, D., Nicholls, G., Bacon, Star, S. J. and Byrom, B. (2004). Optimizing the supply chain through trial simulation. *Applied Clinical Trials*.
- Downer, R. G. (2000). A primer in spatial simulation with PROC SIM2D. SUGI 25 Paper 270-25.
- Duncan, B. (2000). Random simulation methods for computing power and sample size for Tukey's trend test. PharmaSUG, Paper ST10.
- Erdman, D. and Morelock, M. M. (1996). A study of kinetics: the estimation and simulation of systems of first-order differential equations. SUGI 21, Paper 257-21.
- Erdman, D. Simulation models for business planning and economic forecasting. Available at: [support.sas.com/rnd/app/papers/simulationmodels.pdf](http://support.sas.com/rnd/app/papers/simulationmodels.pdf).
- Gastanagaa, V. M., McLaren, C. E., and Delfino, R. J. (2006). Power calculations for generalized linear models in observational longitudinal studies: A simulation approach in SAS. *Comp. Meth. Prog. Biom.*, **84**, 27–33.
- Gemperli, A. (2010). Safety analysis using Bayesian simulation methods in SAS 9.2. *Pharmaceutical Programming*, **3**, 29-36.
- Hamm, A. (2010). Implementing the Continuous Reassessment Method (CRM) in a Phase I Oncology Trial.– Presentation at BASS 2010.
- Han.L. (2006). Generating multivariate normal data by using PROC IML. SESUG, Paper CC15.
- Hanley, J. A. and MacGibbon, B. (2006). Creating non-parametric bootstrap samples using Poisson frequencies. *Comp. Meth. Prog. Biom*, **83**, 57–62.
- Hoffman, D., Kringle, R., Lockwood, G., Turpault, S., Yow, E., and Mathieu, G. (2005). Nonlinear mixed effects modeling for estimation of steady state attainment. *Pharm. Statist.*, **4**, 15-24.
- Hui-Mina, W., Ming-Fang, Y., and Chen, T., H-H. (2004). SAS macro program for non-homogeneous Markov process in modeling multi-state disease progression. *Comp. Meth. Prog. Biom.*, **75**, 95-105.
- Ishizuka, N. (2001). An Application of Continual Reassessment Method Using SAS System. PharmaSUG, Paper SP13.

- Khutoryansky, N. M. (2004). Incremental methods of imputation for longitudinal data with informative missingness. PharmaSUG, Paper SP11.
- Khutoryansky, N.M. (2003). Parameter estimation for incomplete bivariate longitudinal data in clinical trials. PharmaSUG, SP107.
- Khutoryansky, N. M. (2002). Imputation of longitudinal data in clinical trials. PharmaSUG, Paper P10.
- Khutoryansky, N.M. and Huang, W-C. (2001). Imputation techniques using SAS software for incomplete data in diabetes clinical trials. PharmaSUG, Paper SP09.
- Litaker, M. A. and Ferris, D. G. (2004). Simulation Study to Evaluate ANOVA and GEE for Comparing Correlated Proportions With Missing Values. SESUG, Paper SD-06.
- McEntegart, D., Lang, M., Byrom, B., Bacon, S., Star, J., Dowlman, N., and Nicholls, G. (2004). Optimizing the Supply Chain Through Trial Simulation. July 2004.
- Meira-Machado, L., de Uña-Álvarez, J., and Cadarso-Suárez, C. (2009). Multi-state models for the analysis of time-to-event data. *Stat. Meth. Medic. Res.*, **18**, 195–222.
- Miller, D. P. (2004). Bootstrap 101: Obtain robust confidence intervals for any statistic. SUGI 29, Paper 193-29.
- Monti (2004). Stratified Randomization in Clinical Trials. Powerpoint presentation at: [www.rhoworld.com/presentations\\_page/mbc\\_2004.ppt](http://www.rhoworld.com/presentations_page/mbc_2004.ppt).
- Mueller-Cohrs, J. (2006). Analysis of incomplete two-period crossover trials with SAS PROC MIXED. PhUSE, Paper ST07.
- O'Kelly, M. (2009). R vs. SAS in model based drug development. Presentation at 2009 UseR! Conference.
- O'Quigley, J. and Shen, L. Z. (1996). Continual reassessment method: a likelihood approach. *Biometrics*, **52**, 673-684.
- Panbianco, D.L. and Maes, A. (2011). Estimating time to steady state using the effective rate of drug accumulation. *Pharmaceut. Statist.*, **10**, 27–33.
- Parker, I. and O'Kelly, M. (2009). An application of SAS in model based drug development. PhUSE, Paper SP05.
- Patterson, S. C. and Jones, B. (2004). Simulation assessments of statistical aspects of bioequivalence in the pharmaceutical industry. *Pharmaceut. Statist.*, **3**, 13–23.
- Patterson, S. D. and Yeh, S-T. (2007). SAS ® Markov Chain Monte Carlo (MCMC) Simulation in Practice. Paper SP07
- Peszek, I. (2004). Statistical simulations for sample size calculation with PROC IML. PharmaSUG, Paper SP08.
- Rosenkranz, G.K. (2009). Modeling laboratory data from clinical trials. *Comp. Statist. and Data Analysis.*, **53**, 812-819.
- Rosthøj, S., Andersen, P.K., and Abildstrom, S. Z. (2004). SAS macros for estimation of the cumulative incidence functions based on a Cox regression model for competing risks survival data. *Comp. Meth. Prog. Biom.*, **74**, 69-75.
- Russek-Cohen, E., Martinez, M. N., and Nevius, A. B. (2005). A SAS/IML program for simulating pharmacokinetic data. *Comp. Meth. Prog. Biom.*, **78**, 39-60.
- Shen, D. and Lu, Z. (2007). Population pharmacokinetics studies with nonlinear mixed effects modeling. SAS Global Forum, Paper 148-2007.
- Su, J., Sun, P., Li, X., and Hartford, A. H. (2009). Fitting compartmental models to multiple dose pharmacokinetic data using SAS PROC NL MIXED. PharmaSUG, Paper SP11-2009.
- Satya Prasad Tangellamudi, S. P. (2007). A SAS based tool to simulate pharmacogenetic trials, MSc Thesis, Cranfield University.
- Segreti, A. C., Leung, H. M., Koch, G. G., Davis, R. L., Mohberg, N. R., and Peace, K. E. (2001). Biopharmaceutical statistics in a pharmaceutical regulated environment: past, present, and future. *J. Biopharm. Statist.*, **11**, 347–372.
- Smith, M. K. (2003). Software for non-linear mixed effects modelling: a review of several packages. *Pharmaceut. Statist.*, **2**: 69–75.
- Tanriverdio, F. and Ring, A. (2007). Implementation of bootstrapping in a replicative bioequivalence study with a highly variable drug. PhUSE, Paper ST03.

Tavares, A., Paes, A. T., and Carlos, A., and de Lima, P. (2004). A SAS macro for estimating transition probabilities in semiparametric models for recurrent events. *Comp. Meth. Prog. Biom.*, **75**, 59-65.

UMICH (2008). TITE-CRM Phase I Clinical Trials: Implementation Using SAS. University of Michigan Comprehensive Cancer Center Biostatistics Unit.

Wang, J. (2005). A semi-parametric approach to fitting a nonlinear mixed PK/PD model with an effect compartment using SAS. *Pharmaceut. Statist.*, **4**, 59–69.

Wang, J., James R. Carpenter, J. J. R., and Kepler, M. A. (2006). Using SAS to conduct nonparametric residual bootstrap multilevel modeling with a small number of groups. *Comp. Meth. Prog. Biom.*, **82**, 130–143.

Wanga, S., Xiaa, J., Yua, L., Li, C., and Li Xu, L. (2008). A SAS macro for sample size adjustment and randomization test for internal pilot study. *Comp. Meth. Prog. Biom.*, **90**, 66–88.

Welch, B. and Slater, R. (2010). Data simulation for piloting clinical trials display programs. SAS Global Forum, Paper 079-2010.

Westfall, P.H., Tsai, K., Ogenstad, S., Tomoiaga, A., Moseley, S., Lu, Y. (2008). Clinical trials simulation: A statistical approach. *J. Biopharm. Statist.*, **18**, 611–630, 2008.

Westfall, P., Dmitrienko, A., DeSouza, C., and Tomoiaga, A. (2010). Clinical Trials Simulation: A Publicly Available, Grid-Enabled, GUI-Driven SAS®. System. SAS Global Forum, Paper 180-2010.

Whitehead, A., John Whitehead, J., Susan Todd, S., Zhou, Y., and Smith, M. K. (2008). Fitting models for the joint action of two drugs using SAS®. *Pharmaceut. Statist.*, **7**, 272–284.

Whitehead, J. (2010). Statistical group sequential trials revisited: Simple implementation using SAS. *Statist. Meth. Medic. Res.*, Preprint.

Zhang, X. and Zhang, M-J. (2010). SAS macros for estimation of direct adjusted cumulative incidence curves under proportional subdistribution hazards models. *Comp. Meth. Prog. Biom.*, **101**, 87-93.

Ying, G-S. and Heitjan, D.F. (2008). Weibull prediction of event times in clinical trials. *Pharmaceut. Statist.*, **7**, 107–120.

Ying, G-S. (2010). Weibull Prediction of Event Times in Randomized Clinical Trials. Powerpoint presentation – available at: [www.psiweb.org/docs/psijournalclub-ying-2010-10-28.pdf](http://www.psiweb.org/docs/psijournalclub-ying-2010-10-28.pdf).

Zellner, D., Zellner, G. E., and Keller, F. (2001). A SAS macro for sample size re-estimation. *Comp. Meth. Prog. Biom.*, **65**, 83–190.

Zhang, W. (2006). Using the RAND function in SAS® for data simulation in clinical trials. SUG31, Paper 198-31.

Zhou, J., Li, J., Coate, B., and Chen, M. (2006). Empirical power estimation for Phase I dose proportionality studies based on power-law model using confidence interval criteria. SUGI 31, Paper 182-31.

Zhou, J., Yuan, Y., Chen, M., and Coate, B. (2005). Empirical power for higher-order crossover designs in comparative bioavailability clinical trials. SUGI 30, Paper 184-30.

Zhou, J., Coate, B., and He, Y. (2004). Empirical power computation using SAS® for Schuirmann's two one-sided tests procedure in clinical pharmacokinetic drug-drug interaction studies. SUGI 29, Paper 154-29.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Melvin S. Munsaka, PhD  
Senior Manager, Safety Statistics  
Takeda Global Research and Development  
One Takeda Drive  
Deerfield, Illinois 60031  
Work Phone: 224-554-2846  
E-mail: [mmunsaka@tgrd.com](mailto:mmunsaka@tgrd.com)

Michael Carniello, MS  
Senior Manager  
Takeda Global Research and Development  
One Takeda Drive  
Deerfield, Illinois 60031  
Work Phone: 224-554-1926  
E-mail: [mcarniello@tgrd.com](mailto:mcarniello@tgrd.com)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.  
Other brand and product names are trademarks of their respective companies