Paper 180-2010

Clinical Trials Simulation: A Publicly Available, Grid-Enabled, GUI-Driven SAS® System

Peter Westfall, Texas Tech University, Lubbock, TX
Alex Dmitrienko, Eli Lilly and Company, Indianapolis, IN
Cynthia DeSouza, Vertex Pharmaceuticals, Cambridge, MA
Alin Tomoiaga, Texas Tech University, Lubbock, TX

ABSTRACT

A clinical trials system developed for Vertex Pharmaceuticals has become publically available on the Biopharmaceutical Network http://www.biopharmnet.com/innovation/trial_simulation/cts1.php. The system is a SAS/AF® based interface in which clinical trials' simulation parameters can be entered; it can operate either on a SAS grid or as a stand-alone process. The statistical details are published in Westfall P.H., Tsai K., Ogenstad S., Tomoiaga A., Moseley S., and Lu Y. (2008). Clinical Trials Simulation: A Statistical Approach. Journal of Biopharmaceutical Statistics 18, 611-630.

The system provides a generic template for simulations that are typically required to plan clinical trials. Realistic clinical-trial data sets are created using a unifying model that allows general correlation structures for endpoint*timepoint data and nonnormal distributions (including time-to-event), with computationally efficient algorithms. The model allows for patient dropout and noncompliance. An example is given.

INTRODUCTION

As noted in the paper "Clinical Trials Simulation: A Statistical Approach" by Westfall et al (2008), from which much of this introductory section is taken, "drug development is not for the faint hearted." Costs of failed drugs run into the tens or hundreds of millions of dollars, and opportunity costs of failing to bring good drugs to market can be similarly high. At all stages of drug development, "go/no go" decisions must be made with these figures in the balance.

Clinical trials simulation (CTS) helps minimize risks and guide decision making by quantifying and evaluating decisions in the face of uncertainties. The 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. CTS incorporates accessible scientific knowledge to help the entire project team communicate and test ideas, and to plan effective trials for every phase of clinical development. The trial simulation helps the team to anticipate risks and preview the range of expected results before huge investments are committed. An important use of CTS is the development of "mock up" trials: project team members from various disciplines utilize the CTS to explore a series of scenarios, from different trial designs, to alternative ways of analyzing the generated data. As result, the project team can receive prompt feedback on the impact that alternative designs and analysis methods will have on in the future outcomes.

Because trials and data resulting from them are often too complex to allow simple decision-theoretic solutions, interest in CTS has recently exploded in popularity among statisticians, clinicians, and phamacokineticists. Current trends within the pharmaceutical industry and within the offices of some regulatory agencies suggest a promising future for modeling and simulation -- incorporating pharmacokinetic modeling and clinical trial simulation. The support from government authorities has boosted the interest in CTS. Besides the more straightforward application of sample size allocation, design optimization may include protocols (for example, choice of ideal models and test statistics), and estimation of operating characteristics of nonstandard and computationally intensive procedures (including Bayesian and adaptive designs).

Figure 1 displays the essential idea of what we mean when we refer to CTS; others may emphasize different aspects. Often, "variations in study design" refers simply to different sample sizes, but the idea is much broader, encompassing length of trial, measurement of endpoints (continuous, time-to-event, categorized, binary), and analysis methods (baseline covariate-adjusted vs. percentage change, use of compliance data as covariates, parametric vs. nonparametric etc.). The horizontal axis of Figure 1 need not be ordinal; the graph suggests ordinality for illustrative purposes only.

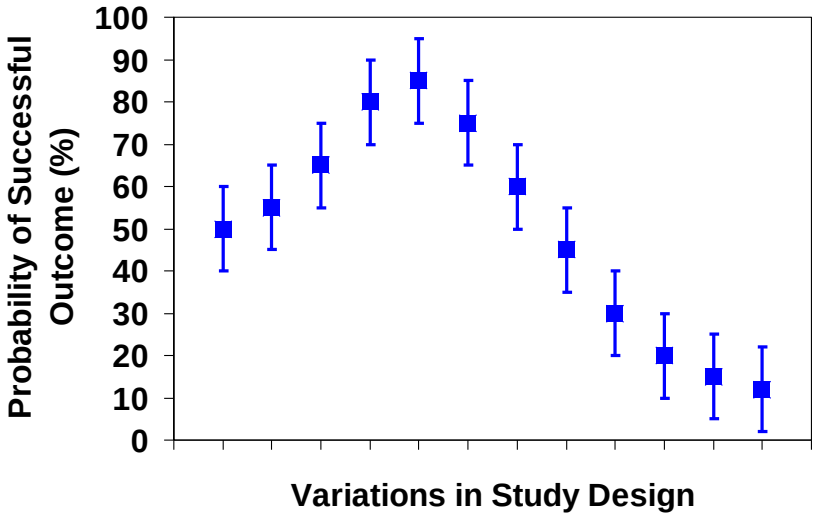

Figure 1. Typical output from a clinical trials simulation system.

Similarly, while "probability of successful outcome" often means "power," the possibilities are much broader, encompassing combination rules involving both safety and efficacy, or complex rules like "3 out of 4 significant" for multiple co-primary endpoints, and rules that include economic considerations, and rules involving patient quality of life. In the more general case, the vertical axis of Figure 1 will be replaced by "Expected Benefit."

EVOLUTION OF SYSTEM DEVELOPMENT: FROM VERTEX TO BIOPHARMNET

In 2005-2006, a SAS system for CTS was developed under the auspices of an agreement between Vertex Pharmaceuticals and Texas Tech University. In 2009, Vertex agreed to make the system publically available under the Biopharmaceutical Network's website. The Biopharmaceutical Network (BioPharmNet.com) is a nonprofit professional organization aimed at promoting collaboration, novel scientific solutions and knowledge sharing in biopharmaceutical drug development. Its mission is twofold, first to facilitate discussion and collaboration, bringing stakeholders from industry, academia and regulatory agencies to a virtual table for discussion of key issues in drug development with emphasis on underlying science and novel methodologies. The second aspect of the mission is to manage knowledge by building a central information resource for scientists involved in drug development to help create and leverage knowledge across the biopharmaceutical industry. The advisory board contains many well-known statisticians in the Biopharmaceutical world: Christy Chuang-Stein, Ralph D'Agostino, Sr., Alex Dmitrienko, Gregory Enas, Michael Krams, Scott Patterson, José Pinheiro, Frank Shen and Brian Wiens.

With the agreement of the advisory board, the clinical trials system is now hosted on the BioPharmNet website. The system development team consists of the four authors of this paper. Users are free to download the software and the documentation, and modifications or improvements are encouraged. The user development team reserves the right to review and either approve or disapprove of the posting of software updates.

EXAMPLES

The following three examples were analyzed using the system, and show a sample of what is possible. The scope of applications is much broader than the small sampling shown here.

SAMPLE SIZE ALLOCATION

Consider an investigation of an arthritis drug, with the binary outcome ACR20 as the primary endpoint. There will be Control, Low, Mid, and High doses. Expectations are that ACR20 response rates are 30%, 50%, 60% and 70%, respectively, and that patient dropout rates are 5%, 10%, 15%, and 20%, respectively in the four dose groups. All comparisons will be made using Chi-Square Dose/placebo tests, using the fixed sequence multiple comparisons method (High dose first, then Mid dose, then Low dose, tested in order until one fails to achieve significance. The total number of patients is 200, and the question is, how to allocate them among the groups?

Elements that make this problem require simulation (rather than analytical results) are (a) the use of Chi-Square tests, whose mathematical distributions are asymptotic rather than exact in finite samples, (b) the dropout issue, and (c) the use of fixed sequence tests, whose power functions depend on joint distributions rather than marginal distributions.

Using the system, with 20,000 simulated clinical trials per design (using the grid implementation), we obtain the following:

Table 1. Power analysis for alternative design configurations.

Design	High Dose	Med Dose	Low Dose
50,50,50,50	.973	.816	.465
101,33,33,33	.966	.800	.448
95,30,35,40	.981	.822	.426
80,40,40,40	.977	.835	.480
80,35,40,45	.985	.837	.452
74,42,42,42	.976	.834	.484

Entries shown in the table are power using the fixed sequence procedure for the various tests. The bottom design seems preferable if the goal is to maximize probability of detecting Low dose significance while maintaining reasonable power for the Mid and High doses. This allocation is familiar for such designs; see e.g. Hochberg and Tamhane (*Multiple Comparison Procedures*, 1987, pp. 164-169).

CHOICE OF TEST

ACR20 is a composite of the seven endpoints Tender Joint Count, Swollen Joint Count, Patient Global Assessment, Investigator Global Assessment, Subject Disability Assessment, Pain, and Erythrocyte Sedimentation Rate. ACR20 = 1 if there is a 20% improvement in the first two endpoints, and a 20% improvement in at least 3 of the remaining 5 endpoints; ACR20=0 otherwise. Rather than use such a crude binary scoring of the data, which loses information and sensitivity, it has been suggested to use the O'Brien test, which more or less combines data from all endpoints in a continuous scale (Anderson, Bolognese, and Felson (2003), "Comparison of Rheumatoid Arthritis Clinical Trial Outcome Measures," *Arthritis and Rheumatism* 48, 3031-3038).

The following table compares powers of the two tests.

Table 2. Power analysis for alternative test statistics.

Design	O'Brien	ACR20
50,50	.60	.41
70,70	.86	.40
100,100	.98	.58

Clearly, the ACR20 test is wasteful in terms of subjects required.

CHOICE OF DESIGN, TEST, AND DURATION OF STUDY

The system produces data with user-specified distributional characteristics, including outlier-prone mixture distributions, useful for modeling patient populations with "outlier" subgroups, or for modeling percentage change data, which are commonly heavy tailed. The usual statistical tests are known to lack power for these studies, so an

investigation is needed to select the appropriate test, in addition to sample size allocations.

A third aspect that can be investigated is length of study. The system uses as input the responses over time (as established e.g., by time-and-dose models from earlier phase studies), so one can investigate the consequences of shorter trial duration. The user input for time response in this trial looks as follows:

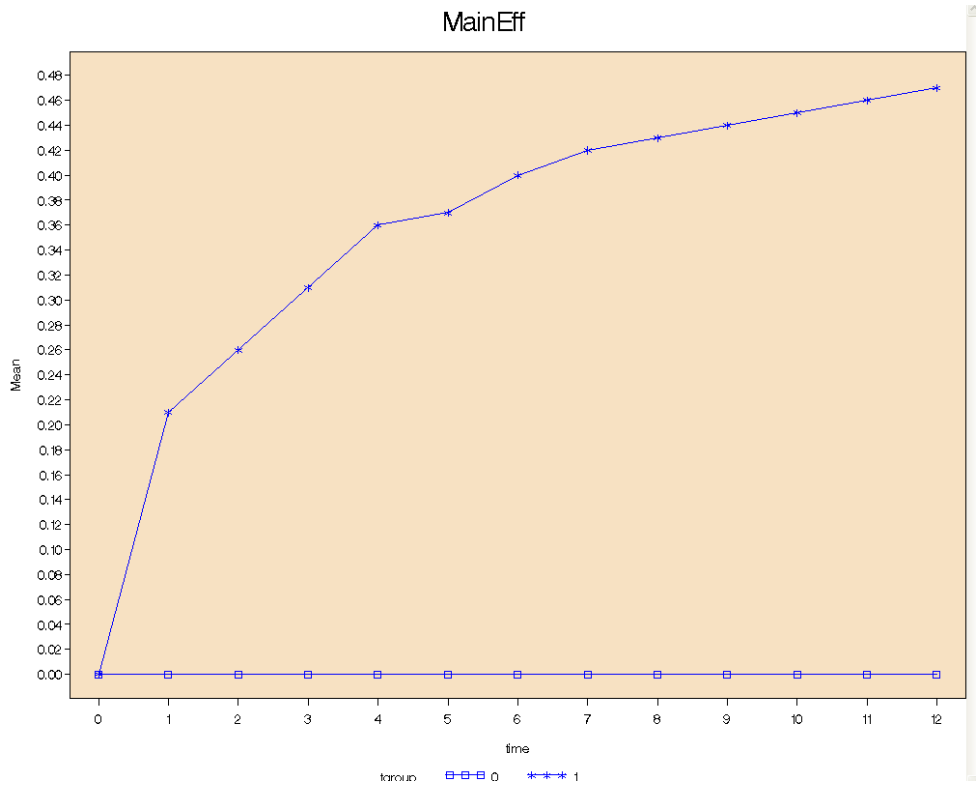

Figure 2. User-input time response functions.

The functions assume a flat placebo response, but a treatment response that increases with time. The measurement standard deviation is assumed to be 1.0, so the effect size is (.47-0)/1 = .47 at 12 weeks and (.43-0)/1 = .43 at 8 weeks.

Dropouts are assumed, and LOCF imputations are used. Based on the inputs, the following simulated power results are obtained for some standard tests, including Kruskal-Wallis (K-W) on raw data and on difference from either baseline median or mean.

Table 3. Power analyses for alternative design, duration, and test statistics.

	Type of Analysis							
Design	AOV	ANCOVA Mean	ANCOVA Median	Difference Mean	Difference Median	K-W	K-W Diff Mean	K-W Diff Median
12 wks/ 30,30	0.41	0.55	0.54	0.48	0.47	0.58	0.67	0.65
12 wks/ 50,50	0.57	0.73	0.72	0.67	0.64	0.80	0.87	0.86
12 wks/ 100,100	0.83	0.94	0.93	0.90	0.89	0.97	0.99	0.99
8 wks/ 30,30	0.36	0.49	0.48	0.43	0.41	0.51	0.59	0.57
8 wks/ 50,50	0.51	0.67	0.66	0.59	0.58	0.73	0.82	0.80
8 wks/ 100,100	0.78	0.90	0.90	0.86	0.84	0.95	0.98	0.98

There are 20,000 simulations per design, all run using the grid. The Kruskal-Wallis test using the difference from the mean of the baseline measurements is best, regardless of design. Obviously, a longer duration will achieve higher power, but one can investigate costs of additional patients versus additional time on trial among trials with adequate (eg 80%) power to determine the optimal combination of number of patients and time on trial. For example, one can investigate further designs to find how many patients are needed per arm in a 12-week study (obviously less than 50 per arm when the K-W difference test is used), and compare costs with the 8 weeks/ 50,50 design.

SYSTEM REQUIREMENTS AND INSTALLATION

The BioPharmNet website is http://biopharmnet.com/ , and the direct link to the clinical trials simulation software is http://www.biopharmnet.com/innovation/trial_simulation/cts1.php. On that page there is a "clinical trial simulation system archive" (direct link http://www.biopharmnet.com/doc/2010_02_13_cts_archive.zip) as well as system documentation (direct link http://www.biopharmnet.com/doc/2010_02_13_cts_documentation.pdf).

The system requires at least a client (or local machine), and optionally, host machines (for grid runs). The system requires SAS/Windows for the client (local) machine with Version 9 or higher, (the system runs with partial functionality under Version 8), including SAS/BASE, SAS/STAT, and SAS/AF for local runs. SAS/GRAPH is desirable as well, but not necessary. For grid runs, SAS/CONNECT is also needed for client and hosts, and SAS/BASE, SAS/STAT are needed on the hosts, but can be in any operating system.

Inside the documentation, instructions are given for installation:

UNPACKING THE FILES

Unzip the archived file to a directory (we will call it INTERFACE_PATH from now on; in the screen shot below, INTERFACE_PATH is C:\Research\TTU_and_CAABI\Vertex\local2_10\local. Do not use a backslash after the path.)

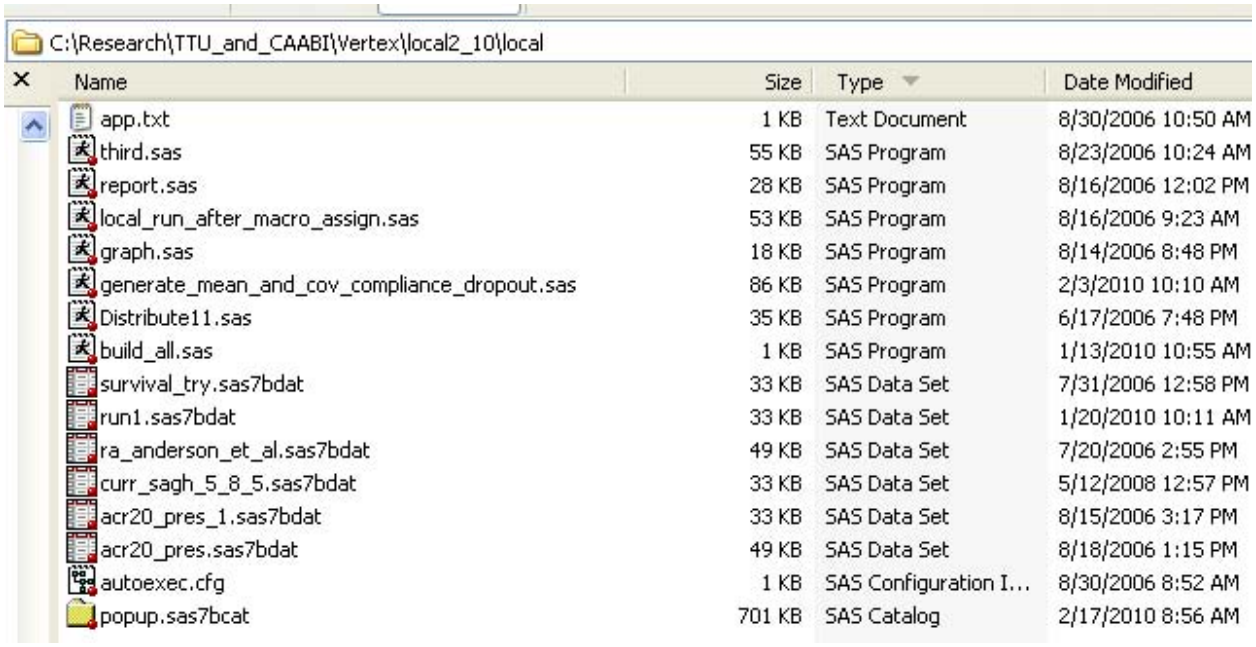

Figure 3. The unpacked files in the local directory.

SETTING THE PATH

Open build_all.sas and set the macro variable path to INTERFACE_PATH.

```
%let path =C:\Research\TTU_and_CAABI\Vertex\local2_10\local;   /* This line  has to be
changed; no space before the semicolon; no backslash at the end. */
```

```
proc build c=popup.popup batch;
```

…

BATCH AND INTERACTIVE STARTING MODES

First, create a SAS library called TRUELOCA that points to INTERFACE_PATH. Then, two separate ways of continuing from this point on are provided:

a. Batch file that automatically starts a SAS session. (All other SAS sessions must be closed in order for the access to the SAS libraries to be unlocked)

a.i. In the INTERFACE_PATH directory there are 2 files that need to be edited:

1. app.bat - sasroot needs to be set to the SAS installation directory (eg, C:\Program Files\SAS\SAS 9.1)

2. autoexec.cfg - the path parameter has to be updated to the location of the clinical trials system application directory (e.g., C:\Research\TTU_and_CAABI\Vertex\local2_10\local)

a.ii. Double click app.bat.

This way of running the application has the advantage that once the a.i. step has been completed once, the user only needs to apply step a.ii.

b. SAS application that requires starting the SAS system (Interactive mode).

b.i. From within a SAS session, run build_all.sas

b.ii. Access the Popup catalog inside the TRUELOCA library; double click it.

RUNNING THE APPLICATION

STARTING THE SYSTEM

(IMPORTANT: Close all the frames and scl files that you might have opened for editing. Save and close all of them before proceeding. Failing to do so, will cause your application not to run correctly.)

If the start is interactive, then run frame F1 by right-clicking the F1 frame and selecting “run,” and frame 1 below appears. If the batch start is used, frame 1 below appears automatically. Here is frame 1:

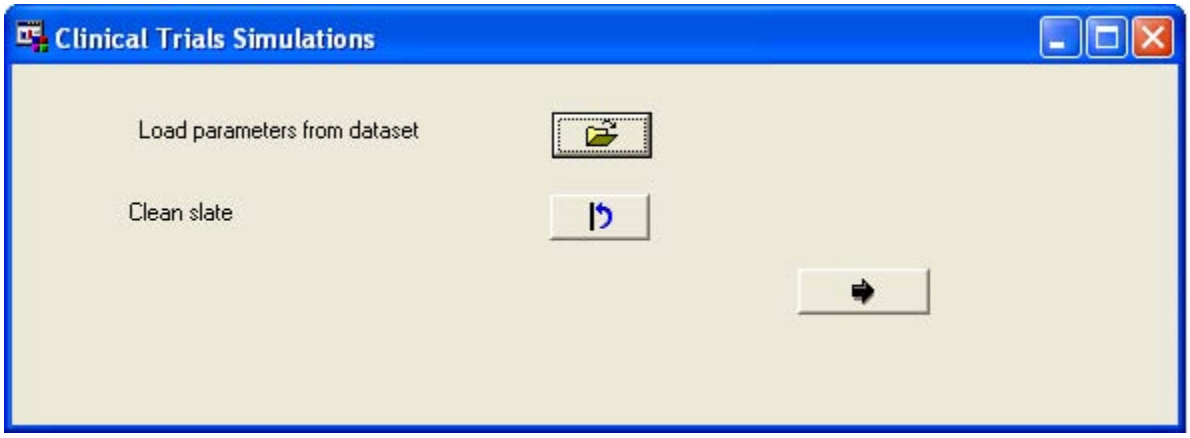

Figure 4. Opening frame from system start.

The first frame is the place for the user to optionally upload an input parameters dataset, saved from a previous run. If the “Load parameters from dataset” button is pressed, the user is going to be presented with a file dialog, allowing

them to choose the desired dataset. If the "Clean slate" button is pressed, there will be no parameters loaded and the user has to input all the parameters of the trial.

A dataset containing the input parameters may be saved after the application has been run, so that the user does not have to re-input all the trial parameters in subsequent runs.

LOCAL OR GRID RUNS

The second frame is the place to choose between a local run (on the local machine) or a grid run. (Remotely, larger simulations can be run in shorter time)

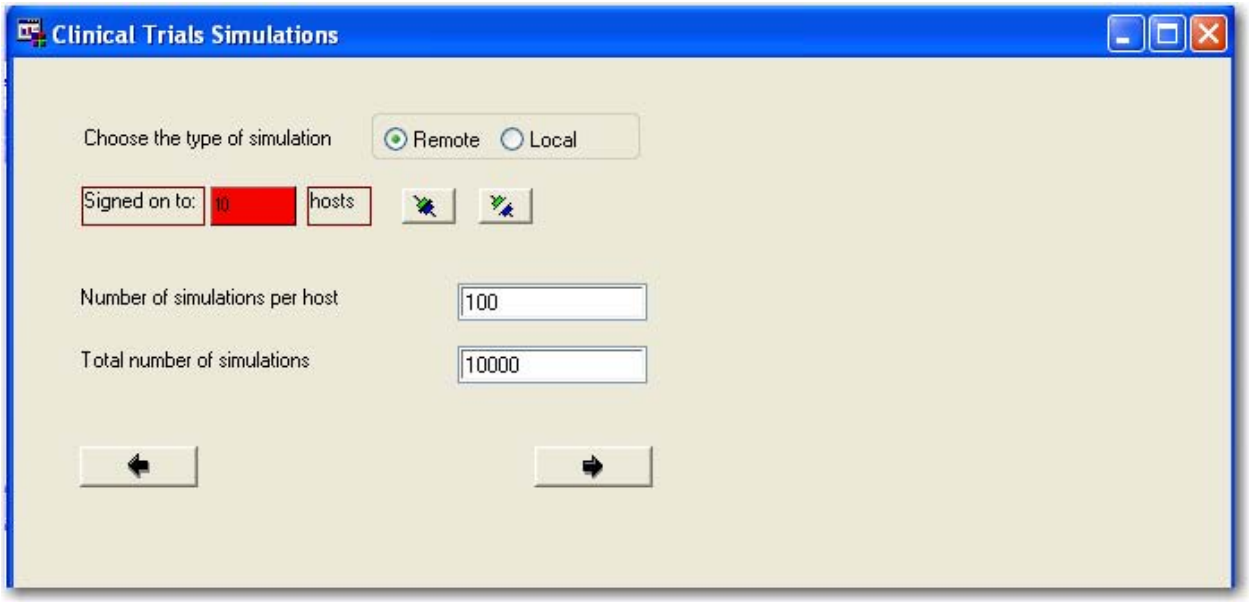

Figure 5. Second frame from system start.

For a remote (grid run), do the following:

a. To run the file on the grid, you must first create (or have available) a SAS data set having the host computer names as well as their username and password logins. For example,

```
data _hosts;
  input host $7. username $10. password $11.;
cards;
bam237 ********* **********
bam238 ********* **********
...
;
```

b. Click the  icon and select the SAS data set containing the host computer names, userids, and passwords.

The "Number of simulations per host" should be large enough so that each run takes several seconds, perhaps as much as half a minute, on the host machines. The "Total number of simulations" should be large enough so that the success measures (usually binary proportions) are estimated with sufficient accuracy; the simulation standard error is given by $(p(1-p)/(\text{Total number of simulations}))^{1/2}$, where p is the success proportion.

Some trial and error is often needed to see which combinations run the fastest. Contributing factors to speed (or lack thereof) are computer latency (smaller numbers of simulations per host mean greater latency) and size of data sets stored on host machines (larger numbers of simulations per host mean larger data sets). When data sets get too large, the time needed to process them increases at a higher rate than linear because there are sort operations within each run.

c. If the user is signed on to the grid, but a remote run is desired using different machines, then the user should sign off using the  icon, then re-sign in using a different machine list by selecting the sign on icon  .

For a local run, only the local computer will be used for the entire computation. In this case only the "Total number of simulations" is needed, and the "Number of simulations per host" is blanked out.

INPUT OF CLINICAL TRIAL PARAMETERS

From this point, the user is asked to provide or modify simulation specific inputs. For details not contained herein, see "Clinical Trials Simulation: A Statistical Approach" by Westfall et al (2008).

Many of the fields may be left blank, and the system with supply default values when the arrow (forward or backward) button is pressed.

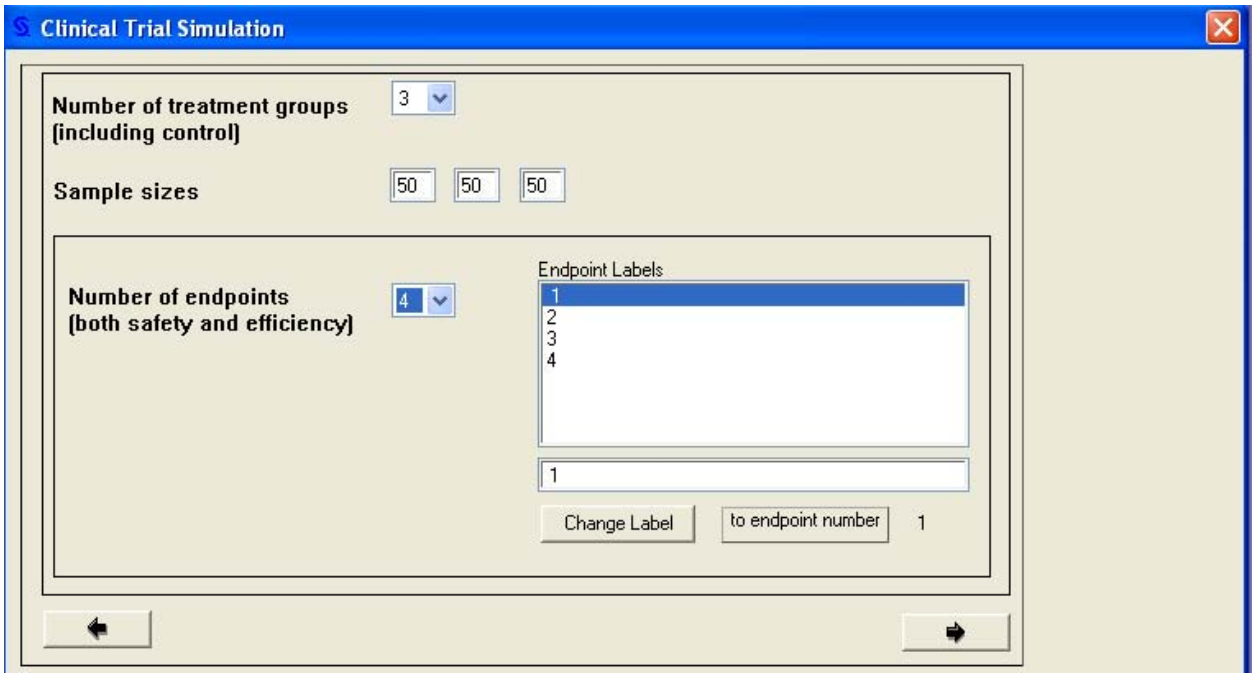

Figure 6. Frame 3, for inputting treatment groups, sample sizes, and endpoint labels.

This frame is largely self-explanatory, but it is worth noting that the endpoints can be given SAS labels in this frame, and future references to the endpoints will use these labels. Otherwise, the labels are generically assigned as "Endpoint1", "Endpoint2" etc.

The next frame allows input of compliance and dropout mechanisms.

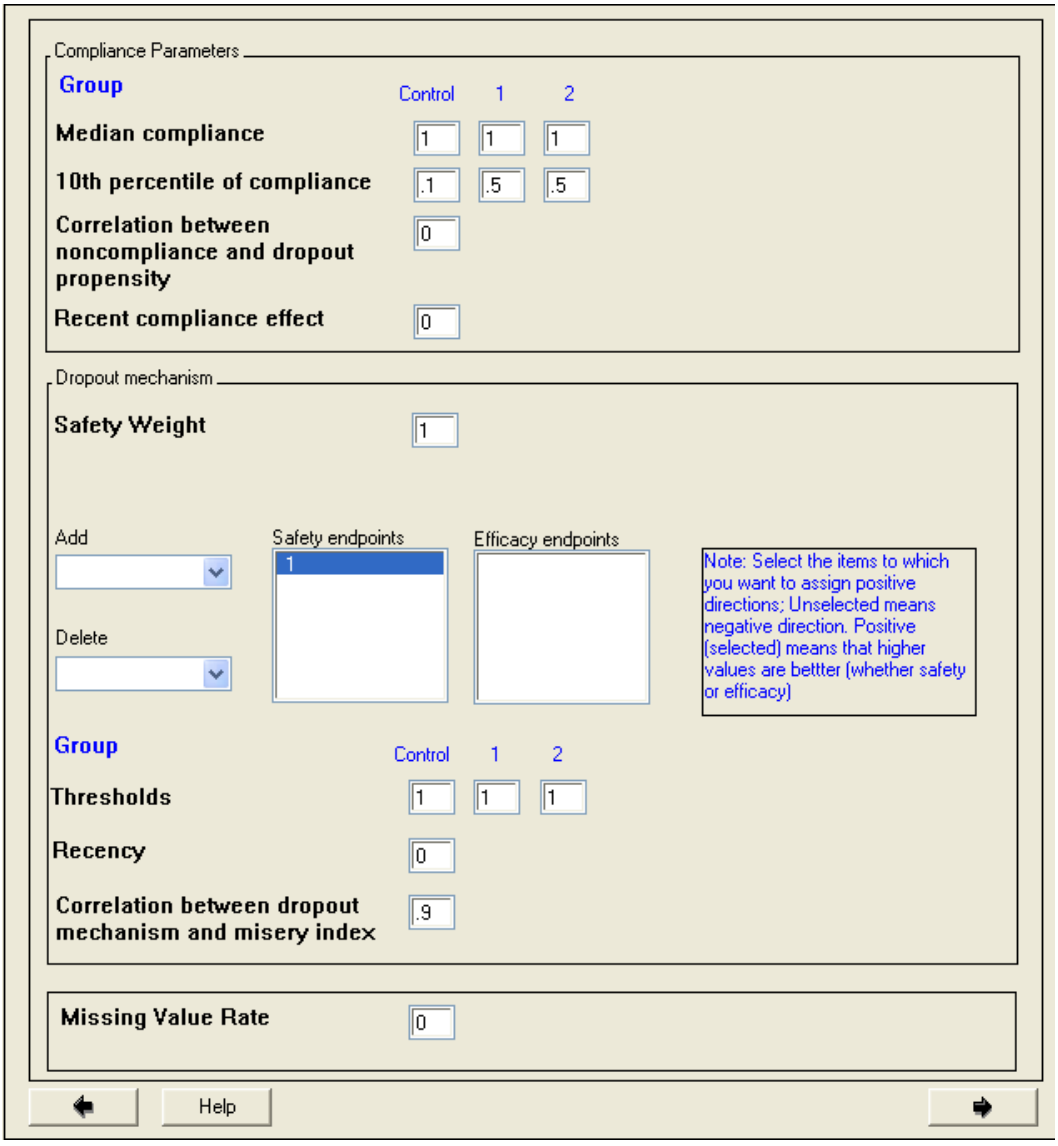

Figure 7. Frame 4, for inputting compliance and dropout mechanisms.

Compliance is assumed to be on the 0 – 1 scale, with 1 denoted perfect, or 100% compliance with the intended dosage. The first inputs are median compliance and 10th percentile of compliance. Using say .9, and .3, respectively, half of the patients are 90% compliant or better, and 90% of the patients are 30% compliant or better. More details are contained in the Westfall et al. report concerning generation of compliance data and effects of noncompliance. If you specify 1 for median compliance, then the system generates perfect (100%) compliance for all patients.

Noncompliance may be related to dropout propensity, if the correlation is high (e.g. 0.9), then noncompliance is closely related to the patient's negative experience in the trial.

The "Recent compliance effect" is a number between 0 and 1 used to determine effect of noncompliance on patient outcome. If this value is "0", then the effect of noncompliance is cumulative over the entire patient history. If the value is "1" the effect is completely determined by the most recent time interval. For values in between 0 and 1, the effect is weighted more heavily by recent compliance using exponential smoothing.

In the Dropout Mechanism box, the user decides which endpoints are involved in the "misery index" that determines patient dropout. The user can choose efficacy and safety endpoints, dropout is a function of both (see Westfall et al, 2008, for details). Dropout rates are controlled by the Threshholds, which determine the proportion of patients staying per visit. The Recency parameter is an exponential smoothing value between 0 and 1 to determine how

persistence of “misery” relates to dropout. If recency is 1, then dropout is determined solely by the most recent weeks experience. If recency is 0, then dropout is determined by cumulative misery. For values between 0 and 1, the more recent history of misery is weighted heavier, again using exponential smoothing.

Since some patients’ dropout is completely independent of their experience in the trial, the software also allows a completely random, non-informative dropout mechanism. The user can input the degree of relation of the misery index with the completely random mechanism in the field “Correlation between dropout mechanism and misery index.”

Finally, there are missing values that differ from dropout. A missing value is simply a missed visit; the patient will visit again in the future. The rate of such missing values can be input in the field “Missing Value Rate.”

Frame 5 allows input of number of timepoints, as well as endpoint and timepoint correlations.

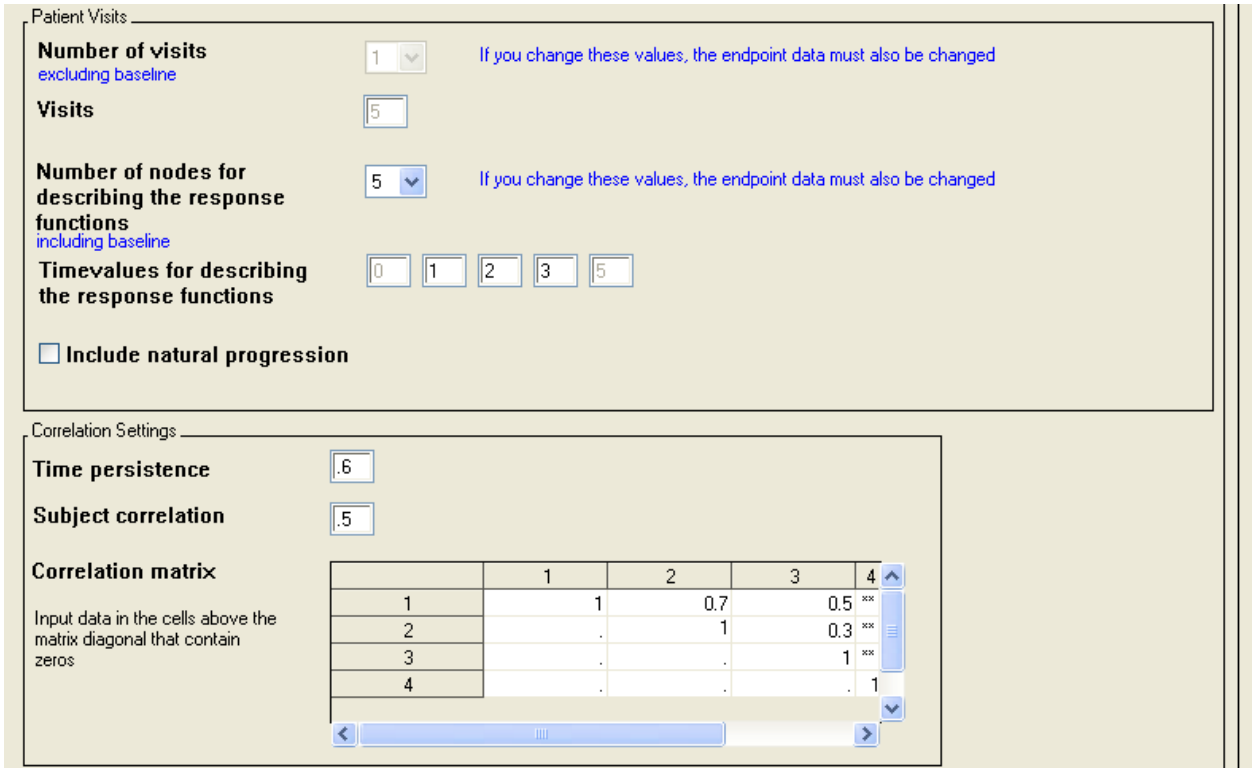

Figure 8. Frame 5, for inputting number of timepoints, as well as endpoint and timepoint correlations.

In this frame, the user specifies how many visits (Number of visits and Visits), and nodes at which to input the time-response functions for each dose (Number of nodes for describing the response functions and Timevalues for describing the response functions). The latter two inputs may be chosen much smaller than the former to save time inputting complex functions if the piecewise linear interpolation used in the software is acceptable for defining time response functions.

The “Include natural progression” field is to be checked if a natural disease history will be input; in this case all patients will regress toward natural history when they are noncompliant. Otherwise they will regress toward placebo.

The “Time Persistence” is the first-order autocorrelation (from an AR(1) process) between time points in the subject-specific model. It must lie between -1 and 1.

The “Subject Correlation” parameter is the intraclass correlation between data values on a single patient for distant time lags (so carryover effects are gone, and all that is left is subject effect). This value lies between 0 (inclusive) and 1 (not inclusive).

The “Correlation matrix” is the correlation between endpoints at a given timepoint. One must double-click the cell in the correlation matrix to enter the values; alternatively, they can be read in from an external data set. A subtle but important detail is that it is the correlation between the underlying normal data, which are essentially latent when the variable’s distribution is chosen to be something other than normal. For example, if the endpoints are binary, then

this matrix contains tetra choric correlations, which are known to be somewhat larger than the correlations between the raw binary data.

Next there are the “Endpoint” specifications. If you select one of them, you get a frame like the following:

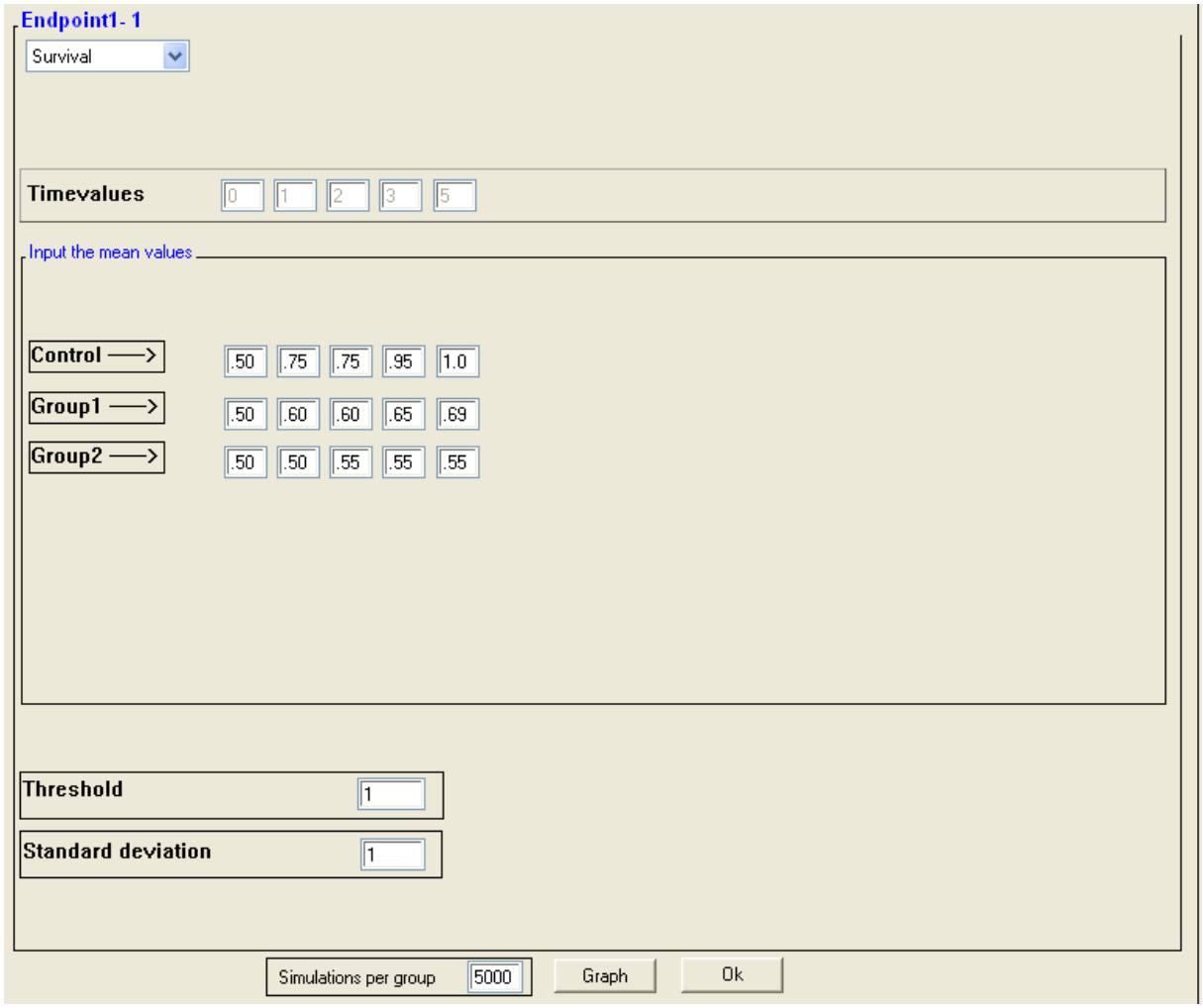

Figure 9. Frame 6, for inputting endpoint specifications.

Here, you can enter the distribution type, as well as the time-and-dose response functions for that variable. There are specific differences corresponding to different distribution types; see the technical report.

After selecting distributions and time-and-dose response functions for all endpoints, the user returns to frame 5. At the bottom there are the final instructions as shown below:

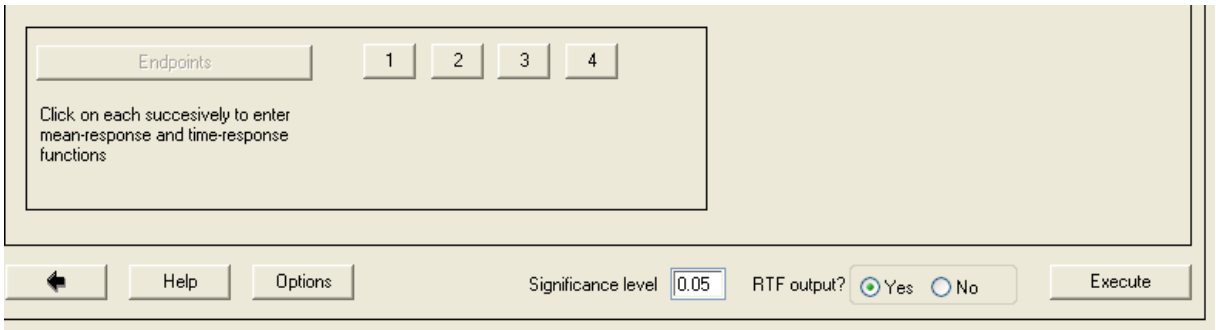

Figure 10. Frame 5’, for final actions.

The “significance level” will be set to the common .05 value by default, but the user can select other values as well. Based on the “Options” selected, the user can either simply “Assign parameters” at this point, and then use the interface for local job runs, or the user can have the system “Execute”, in which case the default analyses will be performed on the generated data sets.

STANDARD STATISTICS COLLECTED

If you run the job locally, the system produces two main data sets: Work.Observed, and Work.Observed_locf. Both are data sets in “multivariate” form, with self-explanatory endpoint and timepoint variable names, along with treatment group indicator and simulation number indicator. The simulation analyses can then be performed using SAS software with “BY” variable processing. For example, one can analyze the data using PROC MIXED to perform longitudinal or Bayesian analyses; or one can use PROC MULTTEST to perform bootstrap and Resampling style analyses, both with the simulation BY variable. The data set “Work.observed” contains all the missing values, whereas “Work.observed_locf” has missing values imputed using “last observation carried forward.”

For grid jobs, the data sets are collated over simulation runs to avoid storage problems. In this case the data sets are over-written, but summary means and p-values are collected and passed back to the client. These data sets are called “Work.sample_means” and “Work-sample”, respectively.

When the “Rtf output” is selected, a report is created after the program execution. This report contains the summary of input parameters selected in the various frames, graphs of means and dropout functions, and summary rejection proportions for the various tests. The variable names are chosen to be reasonably self-explanatory, but can be discerned more specifically from the file “third.sas” which is included in the interface, and which file can be modified to collect customized statistics.

MODIFYING THE SAS/AF GRAPHICAL INTERFACE

The technical report called “clinical trial simulation system documentation” in http://www.biopharmnet.com/innovation/trial_simulation/cts1.php contains details on how one can modify the system to collect other types of statistics and perform analyses other than the defaults provided by the system. Examples inside that documentation show how to allocate sample sizes and to choose statistical tests optimally.

The SUGI paper entitled "Developing SAS/AF(r) Applications Made Easy" is an introduction to SAS/AF Software, and is helpful for understanding more about making these modifications. A copy of this paper can be viewed from the following URL: http://www2.sas.com/proceedings/sugi28/027-28.pdf.

CONCLUSION

The clinical trials simulation system is freely available at http://biopharmnet.com/innovation/trial_simulation/cts1.php. Additional instructions for use with a server are contained therein. Users may freely modify the software to suit their needs. If a particular modification is seen as generally useful, users are encouraged to submit the code and description to the committee comprised of the authors of this paper for inclusion on the website so that others may benefit.

This is the text for the paper’s conclusion.

REFERENCES

Anderson, Bolognese, and Felson (2003), Comparison of Rheumatoid Arthritis Clinical Trial Outcome Measures, *Arthritis and Rheumatism* 48, 3031-3038.

Hochberg, Y. and Tamhane, A. (1987), *Multiple Comparison Procedures*, Wiley, New York.

Westfall P.H., Tsai K., Ogenstad S., Tomoiaga A., Moseley S., and Lu Y. (2008), Clinical Trials Simulation: A Statistical Approach. *Journal of Biopharmaceutical Statistics* 18, 611-630.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Vertex Pharmaceuticals for their financial support of the system, and for allowing other scientists to use it freely. The authors also gratefully acknowledge Kuenhi Tsai, Stephan Ogenstad, Scott Mosely, and Yonggang Lu for their contributions to system development.

CONTACT INFORMATION

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

Name: Peter Westfall
Enterprise: Texas Tech University
Address: Mail Stop 2101
City, State ZIP: Lubbock, Texas 79409-2101
Work Phone: 806-742-2174
Fax: 806-742-3193
E-mail: peter.westfall@ttu.edu
Web: http://westfall.ba.ttu.edu/pwvita.htm