



Tessella and Berry Consultants FACTS™

The Fixed and Adaptive Clinical Trial Simulator

1 Introduction

This document describes the FACTS™ Release 1.1 product from Tessella and Berry Consultants.

FACTS is the result of a strategic collaboration between Tessella Inc and Berry Consultants bringing together the market leading position of each company to produce a powerful tool to help Biostatisticians to rapidly design, compare and simulate both fixed and adaptive trial designs.

For over 25 years, Tessella has been providing quality solutions to a multitude of diverse application areas. Tessella's Clinical Trial Technologies and Services address the unique needs of Adaptive Clinical Trials, allowing the advantages of faster and better-informed decision making in drug development to be fully realized. Beginning with our involvement with Prof. Don Berry and Pfizer in the ASTIN trial, we have continued to be at the forefront of the Adaptive Clinical Trial revolution. Today we can offer products and services to help organizations to fully realize the advantages of faster and better-informed decision making in drug development made possible by Adaptive Clinical Trial Technology.

Berry Consultants have designed and implemented simulators for over 40 adaptive clinical trials. In the last two years, ten of these trials have been run in collaboration with Tessella.

By implementing the design first as a simulator, Berry Consultants allow the design to be evaluated and compared to traditional designs, and they allow the design to be optimized. To meet the demands of simulation, these designs are implemented to execute efficiently. Speed of execution will depend on the complexity of the design and the amount of calculation required, but typically hundreds of simulations can be run in an hour on a conventional laptop.

The designs:

- Fit a selected model to the subject endpoints of interest.
- Based on the properties of the fitted model, evaluate pre-specified decisions. These decisions may include changing the patient allocation probabilities, changing the patient population and/or whether the trial should be stopped for efficacy or futility.



Tessella has collaborated with Berry Consultants in deploying their designs on actual trials, providing validation, hosting of the algorithm, integration with EDC and IVRS systems and automated reports based on the model's output for the DMC.

These simulators are stand alone code modules that accept parameter files and output their results to text files. We refer to these simulators as "Design Engines".

In addition to writing their own Design Engine, Tessella has also created a Graphical User Interface to support Design Engines to allow: the selection of designs, the specification of parameters to those designs, the specification of scenarios to simulate, running simulations and reviewing the results. This software is integrated with grid computing environments (e.g. Condor and Sungrid) so that simulations can be 'off loaded' to a centralized grid resource.

Together these form the Fixed and Adaptive Clinical Trial Simulator: FACTS.

The software tools will support the following key steps in the design of an adaptive clinical trial by enabling simulations to be run to assist with:

- High level decision on the trial design (ability to compare across trial design types)
- Refinement of the trial design to optimize operating characteristics
- Creation of a Simulation Report using output tables exported from trial simulation software

2 Design Engines

There are currently two "Bayesian Dose Finding" Design Engines for Phase 2a/b trials, that analyze two different types of endpoints, one where the endpoint is measured on a continuous scale and one where the endpoint is dichotomous, and a "Dose Escalation" Design Engine for Phase 1/2a trials, that is based on the CRM (Continuous Reassessment Method) and a number of published extensions.

2.1 General Facilities

2.1.1 Patient Response Models

The Design Engines allow responses to be simulated from specified probability distributions

- For continuous endpoints in terms of the variance and the mean response at each dose/treatment arm.
- For dichotomous endpoints in terms of the

probability of response for each arm.

- There are models for simulating subjects' intermediate responses.

Alternatively the user may load a file of simulated subject responses for the simulator to sample from.

2.1.2 Detailed Simulation Model

The simulators include details of trial execution, for instance

- Recruitment rate (average number of subjects to be recruited in a week) is specified per week of the trial. Thus not only can the recruitment rate be varied, but any recruitment profile desired can be specified.
- The time from patient's baseline to response is simulated – including the number and timing of intermediate visits.
- The probability of a subject dropping out on different treatment arms and at different visits can be simulated.

2.2 The Trial Designs

2.2.1 Overall Trial Constraints

The user specifies trial constraints such as

- The maximum number of subjects that can be recruited into the study,
- The different doses of the study drug that can be used
- The target of the trial such as finding the dose with the maximum response, finding a dose that achieves a certain minimum efficacy or maximum tolerability, or finding a particular effective dose quantile.

2.2.2 Modeling alternatives

The user can select from a range of dose response and longitudinal models to fit to the data during the trial, from which posterior probabilities are calculated and trial decisions are based. It is possible to use a 'null model' for comparison purposes.

2.2.3 Dose Allocation Alternatives

In the Bayesian Dose Finding designs the user has options to control the allocation of subjects to different doses:

- to perform no adaptation
- to adapt at fixed interims (using arm dropping)
- to adapt allocation by frequently adjusting the probability of allocation after a 'burn in' period of fixed allocation.



In the Dose Escalation designs the user controls dose escalation by specifying:

- starting dose
- the cohort size
- the number of dose increments from the current dose that can be made in a single step

2.3 Simulation Output

Simulation output is provided at four levels:

1. A high level summary of the simulation results
2. A summary of the result of each simulated trial
3. For the first 100 simulations of any run, the results at each update are output
4. For the first simulation of any run, highly detailed output is produced, including the results for each patient at each visit.

2.4 Design Engine Capabilities for Trial Implementation

The Design Engines include the ability to perform a single model update based on externally provided data and output the corresponding recommendations for the trial. Thus they can be used to execute an actual trial in a manner identical to the designs that have been simulated.

3 Bayesian Dose Finding Designs

3.1 Dose Response Models

There is a single main dose response model, but this model need not apply to all doses. It must be used on a continuous range of doses, but doses at the start and/or end may be excluded from the main model. The models available to model dose-efficacy are:

- Simple NDLM (Normal Dynamic Linear Model)
- Monotonic NDLM
- 2/3 Parameter Logistic Models
- 3 Parameter Logistic with a hierarchical component
- 4 Parameter Sigmoid Model
- No Model

The models are used for Dichotomous Endpoints by

fitting the model to the log odds response.

3.2 Longitudinal Models

By default subjects are observed at a single visit a fixed time after being randomized. Optionally the user may specify one or more earlier visits at which intermediate efficacy observations are obtained.

If there are multiple visits then a longitudinal model is employed to enable final observations to be imputed for those subjects that only have intermediate responses. Different doses may be modeled separately. The following longitudinal models are available.

- Last Observation Carried Forward.
- For Continuous Endpoints:
 - » Simple Linear Regression.
 - » Time Course Hierarchical Model, models the relationship between subjects' early responses and their final response. It differs from the other longitudinal models in that it models all the observations to predict the final value.
 - » Kernel Density, this method is a non-parametric re-sampling approach that is ideal for circumstances where the relationship between the interim time and the final endpoint is not known or not canonical.
- For Dichotomous Endpoints:
 - » A Beta Binomial is used to model the likelihood of a final successful outcome dependent on the result at the interim.

3.3 Allocation

At the start of the trial, fixed numbers of subjects are allocated to each dose, n_d . This is referred to as the burn-in for each dose. The burn-in is a flexible parameter for each dose. This allows a wide range of designs to be employed, such as a top-down approach, or using a grid of doses initially and opening additional doses after the burn-in.

Thereafter the randomization probabilities for some doses may be set to fixed values. The remaining doses are allocated to adaptively.

For doses that are allocated to adaptively, the user may select this to be in proportion to the probability that dose is one of the targets, or it may be weighted by the reduction in variance expected from adding one more subject to that treatment arm.



If any probability is less than a (possibly) user specified minimum threshold, it is set to zero and the remaining probabilities renormalized so that they sum to 1.

Placebo may be allocated to adaptively, with a fixed probability, or using a block randomization scheme.

3.4 Arm Dropping

The user may select to perform adaptation by arm dropping instead of adjusting randomization probabilities. With the option that either there is a total sample size reduction (the future subjects that would have been allocated to the dropped arm are no longer recruited) or the total sample size remains constant (the future subjects that would have been allocated to the dropped arm are divided equally among the remaining arms).

3.5 Early Termination

A wide range of stopping rules are available for decisions to stop for early efficacy or early futility.

Criteria such as the predictive probability of success in phase 3 and the probability of reaching a clinically significant difference are available. Different criteria and thresholds can be specified and combined to form the rule for stopping.

The user can specify a minimum number of subjects that must have been recruited before the stopping rules are applied.

Whether the simulation terminates early or runs to the maximum number of subjects, the model is fitted to the data again after all outstanding final responses have been received and the final outcome determined. Similar to the early stopping rules, the user selects criteria, and specifies probability thresholds for the determination of a futile or successful trial.

3.6 Frequentist Analysis

At the end of each simulated trial the following frequentist values calculated using LOCF (Last Observation Carried Forward) for missing data:

1. Unadjusted dose-placebo comparisons based on a two-sample t-test (continuous) or chi-squared-test (dichotomous)
2. Dunnett-adjusted dose-placebo comparisons based on a two-sample t-test (continuous) or chi-squared-test (dichotomous).

3. Using the general trend test calculate the t-test statistic and p-value using user supplied contrast coefficients.

4 Dose Escalation Designs

This implementation of the CRM is based on 3 published papers [Goodman], [Braun] and [O'Quigley].

The Bivariate Continual Reassessment Method (bCRM) Design Engine can be used in three modes:

1. CRM (toxicity) for Phase 1 trials with binary toxicity endpoints.
2. bCRM for Phase 1-2 trials with binary efficacy endpoints and binary toxicity endpoints.
3. CRM (efficacy) for Phase 2 trials with binary efficacy endpoints.

In the first mode it can be used to model the difference in response in two distinct subject populations.

4.1 The Basic Method

The steps are:

1. The parameter file is read in and the values defined in it stored.
2. The model specifies probability distributions for binary efficacy and/or toxicity responses.
3. A prior distribution is assumed for the model parameters and target levels for $\text{Pr}(\text{toxicity})$ and/or $\text{Pr}(\text{efficacy})$.
4. The model process may be summarised as:
 - a. Assign first cohort to start dose index
 - b. Simulate efficacy and/or toxicity responses
 - c. Calculate posterior distribution of model parameters (using Bayes theorem)
 - d. Determine if trial should continue
 - e. Calculate the next dose, targeting to MTD (stage 1) or MED (stage 2)
 - f. Return to (b)



4.2 Probability Distributions

The bCRM models one or two responses, toxicity and efficacy over a specified range of doses.

Three models are available:

- Logistic with upper and lower asymptotes and one model parameter pre-defined.
- Tanh
- Power

If studying toxicity only, the user can specify that the study will recruit cohorts into 2 separate “samples”, the results to be modelled jointly using the first two models above, but with an additional parameter ‘b’ for the second group. The user may specify either that cohorts of subjects in both samples are randomised together, or all sample 1 cohorts are allocated and then all sample 2 cohorts.

The selection of dose, transition between phases and stopping will be performed independently for the two populations. A trial in one sample may carry on after the trial in the other sample has stopped.

4.3 Dose Allocation

The dose allocation scheme based on that of [Goodman].

The trial will take place in one or two phases depending on whether the trial is studying toxicity only, efficacy only or both:

1. During the first phase allocation will be targeted at finding the maximum dose whose estimated toxicity is at a user specified threshold (MTD).
2. During the second phase allocation will be targeted at finding the lowest dose whose estimated efficacy is at a user defined efficacy threshold (MED), provided this dose has acceptable toxicity.

Allocation begins at the user specified first dose. Subsequent allocation is subject to the rule that allocation may not take place to a dose more than a specified number of levels higher than the current dose at which allocation is taking place. The user may also select to prevent escalation until a specified number of subjects are on the current dose.

The next dose to allocate is calculated from the model curve using the mean of the current estimate of the fitted parameters.

4.3.1 Allocation to Placebo and Maximum Dose

If studying efficacy only the model implements an adjusted allocation scheme.

The highest and lowest doses are given a user defined fixed probability of allocation that may override the allocation to the current selected dose. Each subject in a cohort is randomised separately. A consequence of this scheme is that the subjects in each cohort will not be allocated to the same dose – they may be allocated to the lowest dose, highest dose or selected dose.

4.3.2 Delayed Responses

An optional delay may be specified, in cohorts, between a subject’s recruitment into the simulation and the generation of the simulated responses (efficacy and/or toxicity).

This delay is counted in terms of a number of cohorts recruited between admission and response.

4.4 Early Termination

4.4.1 MTD finding only

The user may specify rules to:

1. Stop after a specified number of subjects have been treated at the MTD (MED).
3. Stop after the width of $(1 - \alpha) * 100\%$ credible interval for MTD (MED) includes a user specified number of dose levels (or less).. For each dose, calculate the probability that it is the MTD (MED). Stop when there is at least one dose for which the estimated probability exceeds a user specified threshold.

The user may specify more than one stopping rule and for the trial to stop only when all the rules have become true, or to stop when any one of the rules has become true.

4.5 Traditional 3 + 3 for comparison

As an option for comparison purposes, instead of the CRM algorithm, the user can select that the simulations be run using the “3 + 3” method. The 3+3 method only investigates toxicity, only works with a cohort size of 3 and only works with one sample population.



5 General

5.1 Execution Speed & Environment

All design engines are implemented as single executable programs in order to provide fast and efficient simulation.

The expected performance is about 1,000 simulations an hour, but the actual performance is very dependent on the size of the trial being simulated – the model used, the number of subjects, number of visits, number of model updates (interims).

Memory and disk requirements are modest and the GUI will run on any hardware configuration that can comfortably run MS Office applications. The Design Engines too have low memory and disk requirements, when simulating they are computationally intensive, and the times they take to run are directly proportional to the processor speed.

To facilitate running large numbers of simulations, the GUI includes the facility to off load the simulations onto a computing grid. It will write the job details to a shared drive that is monitored by a 'sweeper' process that handles the submission of the job to the grid, thus grid software does not need to be installed on the user's PC or laptop.

5.2 Future Plans

Tessella and Berry Consultants are already planning follow-up releases of FACTS. The content and functionality that will be included in these releases will be strongly influenced by feedback from customers, but is likely to include developments in the following areas:

- Further Design Engines from Berry Consultants, including a Bayesian Dose Finding Design Engine for Time-to-Event endpoints and a Dose Finding Design Engine using a Clinical Utility Index to allow adaptation to be driven by multiple endpoints.
- Extended user interface, particularly in the presentation and analysis of simulation results.

Tessella Inc 233 Needham Street, Suite 300, Newton, MA 02464, USA
T: 1 617 454 1220 | E: info@tessella.com

Tessella plc 26 The Quadrant, Abingdon Science Park, Abingdon, Oxfordshire OX14 3YS, UK
T: +44 (0)1235 555511 | F: +44 (0)1235 553301 | E: info@tessella.com

Tessella – successfully delivering IT and consulting services to world leaders in R&D, science and engineering.

For decades, Tessella has been successfully delivering IT and consulting services to world leaders in R&D, science, and engineering. Through the application of scientific methods and rigorous quality procedures, we enable clients in life sciences, energy, the public sector, and consumer industries to achieve a wide range of objectives, including, forecasting floods, developing fusion power, enhancing military sensor capability, improving drug discovery and development efficiency, and reducing risk to health and the environment in the extraction and production of oil and gas. With offices in Europe and North America, global companies rely on Tessella for business critical assignments.

Copyright © Tessella plc 2009, all trademarks acknowledged. Issue: V1.R0.M0 | Apr 09



www.tessella.com

