Title: Large-scale simulations on a Linux cluster

Authors: (1) David A. James*; (2) David Ohlssen

Institutions: (1) Novartis, East Hanover, NJ, USA; (2) Novartis, East Hanover, NJ, USA.

Objectives: Investigation of operating characteristics of a Bayesian-based futility rule; development of a framework for distributed computing for modeling and simulation.

Large-scale simulations provide an effective means to quantify model variation and uncertainty under plausible scenarios. Some techniques, such as the bootstrap and Bayesian models, may require large number of potentially expensive runs to properly quantify model precision, model uncertainty, etc. Fortunately, many of these problems are known to be "embarrassingly parallel", i.e., they are amenable to be split into many smaller tasks that can run independently of each other on multiple machines.

In this paper, we describe our experience and tools we have developed for running large scale simulations to characterize the operating characteristics of a Bayesian-based futility rule. The rule addresses futility for an endpoint consisting of bivariate over-dispersed Poisson counts, in particular, lack of significant change in the underlying Poisson rates under a treatment arm.

Methods: Bayesian modeling, simulation, distributed computing.

Our modeling and simulation approach consists of three elements: A Bayesian model to account for the bivariate over-dispersed Poisson counts, a futility rule expressed in terms of the Poisson rates estimated from this model, and a relatively large number of simulation scenarios where we investigate the operating characteristics of the futility rule.

Computationally our approach consists of the development of WinBUGS code for Bayesian estimation and R code for simulating Poisson-type of counts according to the various scenarios, and the development of an infrastructure where compute nodes in a Linux cluster can run R natively on Linux and WinBUGS on top of a Windows emulator.

Results: Significantly reduce the time required to simulate all scenarios under study from a few months of CPU time to a few days. This time-savings allowed the team to explore other futility rules as well as sample size and power considerations.

Conclusions: We have described a simple, yet powerful approach to solving one “embarrassingly parallel” type of application, but further work needs to be done to create more general purpose tools for distributed computing for modeling and simulation.

References:

[1] David Spiegelhalter, Andrew Thomas, Nicky Best, Dave Lunn (2003). WinBUGS User Manual. Cambdridge, UK.

[2] R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org..