Performance Characteristics of NONMEM Parallelization

Evan Wang\(^1\), Michael Heathman\(^1\)

\(^1\) Eli Lilly and Company, Indianapolis, IN

**Objectives:** To understand the dependence of NONMEM runtime on the number of cores used and dataset size, for models of varying complexity. This can be used to determine the optimum number of cores for the fastest runtime or highest computational efficiency.

**Methods:** Benchmark runtimes and literature \(^{[1]}\) were used to build a model for NONMEM runtime as a function of the number of cores used and dataset size. To evaluate the ability of the model to fit and ultimately predict NONMEM runtimes, benchmark NONMEM estimations were performed using a “simple” two-compartment pharmacokinetic model (FOCEI) and a “complex” one-target quasi-steady-state TMDD model \(^{[2]}\) (FOCEI, IMPMAP, SAEM), representing the two extremes of runtime performance and efficiency. A series of dataset sizes and number of cores (up to 96 cores) was used. All data were simulated in order to ensure dataset homogeneity. All NONMEM models were run multiple times to average out the effects of network traffic.

**Results:** Predictions for the simple and complex models clearly demonstrate that the runtime model is flexible enough to capture the runtime dependence on cores and dataset size, including a non-monotonic runtime profile. An analytical solution for the number of cores corresponding to the minimum runtime was also derived using the model. With a few benchmark runs, users can apply this model to (1) map out the relationship between the number of cores and runtime, relative speedup, and computational efficiency and (2) find the number of cores that will result in the fastest runtime or highest efficiency.

**Conclusions:** For computationally expensive models, runtime profiling to determine the optimum number of cores is a worthwhile exercise, particularly in cases where a large number of models will need to be run.

**References:**